PRECISE4Q

PREDICTIVE MODELLING IN STROKE

# DELIVERABLE

Project Acronym: **Precise4Q**

Grant Agreement number: **777107**

Project Title: **Personalised Medicine by Predictive Modelling in Stroke for better Quality of Life**

## D1.7 – Report on the ethics of personalized medicine/data-driven multi-dimensional modelling with best practices

Revision: 0.1


| | | | | |
|---|---|---|---|---|
| **Authors and Contributors** | Julia Amann (ETH Zurich), Kelly Ormond (ETH Zurich), Alessandro Blasimme (ETH Zurich), Effy Vayena (ETH Zurich) | | | |
| **Responsible Author** | Julia Amann | | **Email** | julia.amann@hest.ethz.ch |
| | **Beneficiary** | ETH Zurich | **Phone** | +41 44 632 46 19 |



| | | |
|---|---|---|
| Project co-funded by the European Commission within **H2020-SC1-2016-2017/SC1-PM-17-2017** | | |
| Dissemination Level | | |
| PU | Public, fully open | |
| CO | Confidential, restricted under conditions set out in Model Grant Agreement | |
| CI | Classified, information as referred to in Commission Decision 2001/844/EC | |

# Revision History, Status, Abstract, Keywords, Statement of Originality

**Revision History**

| Revision | Date | Author | Organisation | Description |
|---|---|---|---|---|
| 0.1 | 01.11.2021 | J. Amann | ETH Zurich | Preliminary outline |
| 0.2 | 20.12.2021 | J. Amann | ETH Zurich | First draft |
| 0.3 | 07.01.2021 | J. Amann | ETH Zurich | Feedback EV, KO, AB |
| 0.4 | 10.01.2021 | J. Amann | ETH Zurich | Sent for internal review CUB, EMP |
| 0.5 | 11.01.2022 | J. Amann | ETH Zurich | Overview presentation to consortium |
| 0.6 | 17.01.2021 | D. Frey | CUB | Revision |
| 1.0 | 21.01.2022 | J. Amann | ETH Zurich | Submission |
| | | | | |
| | | | | |

| Date of delivery | Contractual: | | Actual: | |
|---|---|---|---|---|
| Status | final ☒ /draft ☐ | | | |

| | |
|---|---|
| Abstract<br>(for dissemination) | This document describes the activities conducted in relation to T1.5, T1.6, and T1.7. These activities contribute to achieving objective 5-7 laid out in WP1 (Patients' Needs and Ethical Framework). More specifically, it describes the rationale, overall methodological approach, and results of these activities. In doing so it builds on and extends previous deliverables (D1.5 and D1.8). The deliverable concludes with a reflective framework to ensure that the PRECISE4Q tools are reconcilable with core ethical values that should guide all clinical decision-making. |
| Keywords | Ethics, AI, autonomy, justice, data protection and privacy, disclosure, responsibility, empathy, explainability |

**Statement of originality**

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

# Table of Contents

# Executive Summary

The overall aim of PRECISE4Q is to minimize the burden of stroke for individuals and society through multi-dimensional predictive modeling. The work presented here contributes to this goal by providing a multifaceted account of the ethics of data-driven multi-dimensional modelling. More specifically, it presents ethical guidelines for data-driven predictive modeling in stroke medicine that are rooted in normative considerations, build on existing ethical frameworks, and consider the lived experience, attitudes, values, and expectations of prospective users, beneficiaries, and developers. This deliverable constitutes a cornerstone of the project when it comes to implementing the PRECISE4Q tools into clinical practice.

The present deliverable outlines activities carried out in relation to T1.5, T1.6, and T1.7 (WP1). More specifically, it describes the rationale, overall methodological approach, and results of these activities. In doing so it builds on and extends previous deliverables (D1.4, D1.5, D1.8, D2.8). The deliverable concludes with a reflective framework to ensure that the PRECISE4Q tools are reconcilable with core ethical values that should guide all clinical decision-making.

Based on the insights gathered through the various stages of the project, we conceptualized a reflective ethical framework consisting of ten sub-sections across development and deployment. The reflective framework aims to guide the final stages of development before bringing the PRECISE4Q tools to market and outlines aspects to consider beyond initial deployment (i.e., continuous monitoring and evaluation). Its main purpose is to stimulate discussion and reflection among the consortium partners, allowing them to identify and anticipate potential ethical challenges that might jeopardize the successful translation of the Precise4Q tools into clinical practice. In doing so, the framework aims to ensure that decision-making during the development and deployment phase is closely aligned with core ethical values and principles of patient-centered care. The framework does not claim to be exhaustive, instead its guiding questions should serve as a basis for interprofessional exchange and reflection within and beyond the consortium, involving also prospective end-users and beneficiaries to provoke an in-depth engagement which may lead to the identification of new questions to be discussed.

# 1 Overall objective and scope of the deliverable

The overall objective of this deliverable is to put forward a reflective framework for data-driven medical prediction in stroke to ensure that the PRECISE4Q tools are reconcilable with core ethical values that should guide all clinical decision-making. The reflective framework presented here is rooted in normative considerations, builds on existing ethical frameworks, but also takes the lived experience, attitudes, values, and expectations of prospective end-users (i.e., clinicians) and beneficiaries (i.e., patients and their families) of AI-powered clinical decision support systems into account. It is also informed by D1.4 (Set of functional requirements and architecture) and D2.8 (Pilot for clinical decision support system). Moreover, we also aimed to actively involve medical researchers and data scientists to also incorporate their views and perspectives.

The present deliverable outlines activities carried out in relation to T1.5, T1.6, and T1.7 (WP1), some of which are currently still ongoing:

- **T1.5 State-of-the-art analysis and empirical study on patients' and clinicians' attitudes towards personalized medicine and multi-dimensional predictive models (M7-M20) ETH,** CUB, QMENTA, UOT *Input: T1.1 - Output: D1.6,* ***D1.7***

- **T1.6 Provide ethical guidelines for data-driven multi-dimensional modelling with the release of a deliberative dashboard to anticipate, discuss and resolve novel ethical issues (M23 – M36) ETH** *Input: T1.1, T1.3, T1.5 / Output:* ***D1.7****, D1.8*

- **T1.7 Assessment of the ethical framework for big data health research, and of the ethical guidelines for data-driven medical prediction in internal surveys and a public symposium (M37-48) ETH,** CUB, EMP
- *Input: : T1.1, T1.3, T1.5 Output: D1.5, D1.6,* ***D1.7***

What follows describes the rationale, overall methodological approach, and results of these activities. In doing so, this deliverable builds on and extends previous deliverables (D1.4, D1.5, D1.8, D2.8).

# 2 Impact of Covid-19 on AI ethics and the Precise4Q ethical framework

Ethical guidance cannot be created in a vacuum but needs to be constantly assessed and evaluated against contextual factors and the broader societal discourse around norms and values. Over the course of the past two years, numerous ethical issues regarding the ongoing Covid-19 pandemic and its impact on healthcare globally have been raised in the academic, policy, and public discourse alike. These considerations have significantly shaped the work presented here.

The Covoid-19 pandemic has claimed many lives and has shaken our society to its very foundations. Healthcare systems globally are collapsing and the long-term consequences for population health and well-being, as well as the economic impact of the crises remain to be seen in the years to come.

The pandemic certainly has a lasting impact on the bioethics landscape and stirred up many questions about the ethical use of big data and AI in times of a public health crises [1-5]. Who should be treated if resources are limited and can artificial intelligence support decision-making process? To what extent can restrictions imposed on individuals or groups of individuals be legitimized in the interest of public health? And what role do digital health technologies powered by artificial intelligence play in this context?

As a society, we are tasked to weigh the potential risks against the potential benefits of introducing novel technologies and facilitating data sharing at the costs of lowering validation standards or disregarding informed consent procedures [6]. But who should decide when it is necessary and in the interest of patients to bring tools to the clinic despite them being not rigorously validated and tested? What are the risks of doing so for patients and healthcare staff? And who bears this responsibility? Similarly, we must ask ourselves, if our efforts to make data more accessible for public health benefits may have opened the door to privacy infringement and abuse by third party actors or totalitarian systems.

# 3 Methodology

We pursued a multi-stage participatory approach to achieve the overall objective of this deliverable (Figure 1). By following a participatory research methodology, we aimed to ensure that the resulting framework is responsive to different stakeholders needs and suitable to guide decision-making in practice, allowing for a thorough and comprehensive assessment of the risks, benefits, and potential pitfalls associated with the Precise4Q tools.

Rooting our analytical approach in the paradigm of patient-centered care [7], we conducted a qualitative study based on semi-structured interviews with stroke survivors, family members, and healthcare professionals specialized in stroke. This study informed the development of a project-internal web-based deliberative dashboard which aimed to foster exchange about pertinent ethical issues and concerns within the consortium (launched in April 2020). In April 2021, we then held an internal impact workshop series to engage consortium partners in a joint discussion about the ethics of predictive modelling in stroke. Our activities were further informed by continuous literature and policy monitoring as well as exchange with other European initiatives and projects that are active in the field of digital health ethics and AI ethics.

In addition to our participation in international events on digital health ethics and ethics of AI, we are in constant exchange with other initiates to learn about best practice approaches in the field. Most notably, Julia Amann is an active contributor to the Z-Inspection® initiative, which pioneers a methodology for the interdisciplinary assessment of trustworthy AI based on ethics guidelines for trustworthy AI put forward by the High-Level Expert Group on AI (AI HLEG) in April 2019 [8, 9]. Members of the Z-Inspection® initiative include researchers and practitioners from medicine, computer science, and social sciences primarily from Europe and North America. Use cases included the assessment of a deep learning based skin lesion classifier [10], a machine learning based tool to assess cardiac arrest in emergency calls [11] and most recently, deep learning for predicting a multi-regional score conveying the degree of lung damage in COVID-19 patients [6].

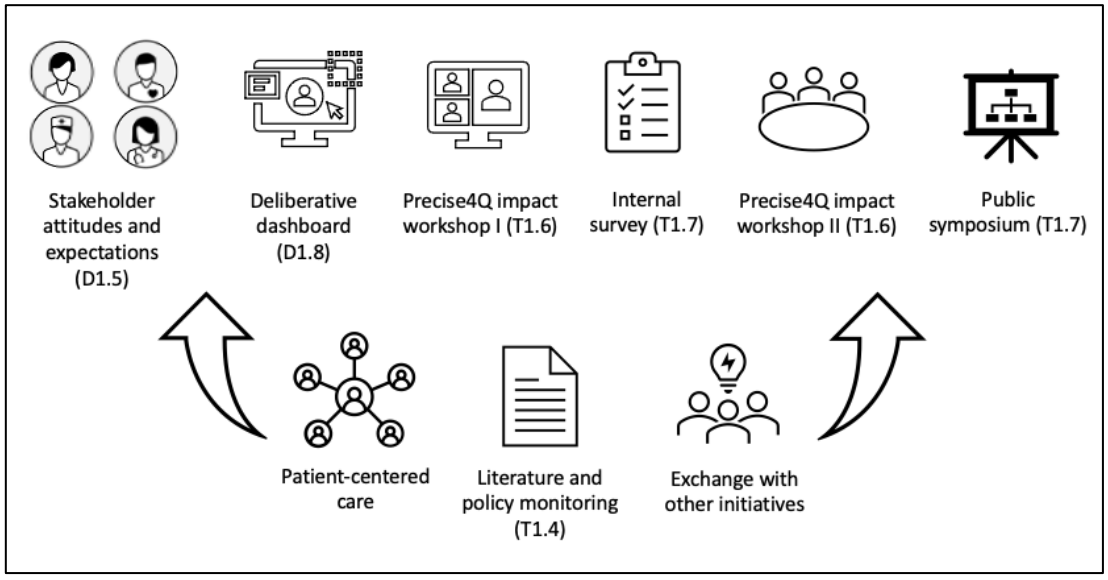



**Figure 1 Methodological approach**

# 4 Literature review

We reviewed the pertinent literature and identified three areas of challenges to consider when seeking to maximize the advantages of multi-dimensional predictive modelling in stroke medicine, and healthcare more generally. Specifically, we identified technological, methodological, and ethical challenges at the stages of 1) data sourcing; 2) application development; and 3) deployment in clinical practice.

We have summarized the results from our literature review in a book chapter, and it has also informed two additional Open Access publications that focus on the ethical considerations of explainable artificial intelligence in clinical decision support systems (CDSS).

Amann J. *(forthcoming)* Machine learning in stroke medicine: Opportunities and challenges for risk prediction and prevention. In F. Jotternad, M. Ienca (Eds.) Artificial Intelligence in Brain and Mental Health: Springer, Europe.

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V.I. (2020) Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Medical Informatics and Decision Making, 20(1):1-9.

Amann J*, Vetter D*, Blomberg SN, Christensen HC, Coffe M, Gerke S, Gilbert TK, Hagendorff T, Holm S, Livne M, Spezzatti A, Strümke I, Zicari, RV Madai, V.I.* *(forthcoming)* To explain or not to explain? - A Case Study of Artificial Intelligence Explainability in Clinical Decision Support Systems. Plos Digital Health.

# 5 Empirical study on attitudes towards predictive modelling in stroke

As part of T1.5 we carried out a qualitative study to explore how stroke survivors, family members, and healthcare professionals specialized in stroke conceptualize AI and how they view the use of AI-powered clinical decision support systems in stroke medicine with a particular focus on the perceived benefits, risks, and ethically relevant concerns.

In continuation of T1.5, we conducted additional interviews in early 2020. In total, we conducted 34 interviews with stroke survivors (N=14), clinicians (N=14), and family members (N=6) which lasted between 22 and 78 minutes. On average, interviews lasted 40 minutes. This led to 22 hours and 38 minutes of interview material which was audio-recorded and transcribed verbatim for analysis. Data was analyzed using a combination of inductive and deductive thematic analysis.

Three core themes emerged from this analysis:

**Presumed roles of AI in the clinical setting**

The theme *Presumed roles of AI in the clinical setting* captured participants conceptualizations of how AI-based systems may be used in the clinical setting. From participants' descriptions and accounts of what they expected an AI system to look like, we identified four potential roles on a continuum raging from very basic to more sophisticated forms of application:

a) AI as an administrative assistant taking over mundane and administrative tasks (e.g., data collection, data synthesis)
b) AI as clinical-decision support, taking on a more assistive role in the decision making process
c) AI as the healthcare professional's right hand, suggesting a more advisory role at eye level with the clinician
d) a fully autonomous AI system that operates fairly independent of the clinician with less opportunity for human intervention.

Understanding how prospective users and beneficiaries view the role of AI in practice is instrumental as it is indicative of how individuals intend to use or expect these systems to be used in practice. This, in turn, has important implications for the type of ethical considerations that should guide the implementation and monitoring of AI-based tools.

**AI offers opportunities but it is not a panacea**

The second theme we identified was *AI offers opportunities, but it is not a panacea*, a theme reflecting both hope and skepticism towards AI-based tools in healthcare. Specifically, we found that despite the perceived opportunities and benefits (i.e., efficiency, quality improvement) participants considered, they also raised concerns that AI-based CDSS may, in fact, not be able to address some of the "most basic" problems in healthcare, or worse, create new problems (further specified in *Perceived challenges, risks, and open questions*). There was almost a uniform agreement that the availability and accessibility of data in the four phases of stroke (prevention, acute, rehabilitation, reintegration) was key to improving healthcare, independent of the use of AI.

**Perceived challenges, risks, and open questions**

The third theme captured participants' considerations regarding *Perceived challenges, risks, and open questions*. A core aspect related to the importance of relational aspects, i.e., the therapeutic alliance. Several participants underlined the need for human touch and empathy in stroke medicine and healthcare more generally, which, in their view, could not be replaced by an AI-based CDSS. The theme also captured the perceived need for human control over decision-making and concerns regarding the potential overreliance on AI-systems, which, according to some of the participants, may in the long run lead to loss of skills and expertise. Finally, the theme touched upon considerations related to patient autonomy, data protection, and privacy, with many participants explicitly stating that patients should be able to make decisions about their data being collected and the options which are informed by this data. There was no common agreement on whether patients would need to be informed by clinicians that there was an AI-based system involved in the decision-making process. While some participants voiced concerns regarding data protection and privacy, there were also does with little or no concerns. Our findings further indicate that AI-based CDSS did not raise any particular privacy concerns other than those of data sharing, more generally.

A detailed description of the methodological approach of the qualitative study has been presented in D1.5 (submitted in Dec 2019). A related publication is currently in preparation and will be submitted shortly.

# 6 Deliberative dashboard

In April 2020, we launched a deliberative dashboard (D1.8, submitted in April 2020), which was envisioned as an interactive platform for consortium partners to engage in joint discussions on ethical issues that may arise within the project. In addition to the interactive forum space, the deliberative dashboard also features a curated, searchable list of ethics resources, including scientific publications, policy documents, videos, and more. The selection of resources was informed by findings of the qualitative study on clinicians', patients', and caregivers' attitudes towards predictive modeling in stroke (D1.5), the categorized and ranked clinical challenges and needs (D1.2) and the use cases and their inputs/outputs specifications (D1.3). Each consortium partner received personalized login information (i.e. user name and password) to access the deliberative dashboard. A detailed description of the deliberative dashboard has been presented in D1.8.

Despite our hope that the deliberative dashboard would foster discussion and reflection on the ethical issues of predictive modelling in stroke medicine, it has been only rarely accessed by the consortium partners. The insights we could gain from the platform were thus limited to the feedback we received from partners bilaterally during the development phase. Consequently, we could not carry out an analysis of the forum content and platform usage (i.e., use of resources), as we had initially planned.

We partly attribute the lack of engagement with the deliberative dashboard to the impact of the pandemic on the overall project and consortium partners' prioritization during this time period. As our attempts to promote uptake were unsuccessful, we took countermeasures to ensure an ongoing interdisciplinary exchange among the consortium partners about the ethical issues related to data-driven predictive modeling in stroke. Specifically, to compensate for the lack of engagement, we launched the Precise4Q impact workshop series (see 7).

# 7 Precise4Q impact workshops

An internal workshop series was held virtually via Zoom in April 2021, comprising of three sessions with five to seven participants each. In addition to Zoom, a virtual whiteboard was set up to facilitate and make collaboration more engaging. Almost all consortium partners had at least one representative attend one of the workshop sessions. Each workshop lasted 2 hours and 45 minutes and combined individual and small group activities with plenary discussion.

The central aims of the workshop series were

1) to refine the consortiums' vision for the Precise4Q tools by generating a more nuanced understanding of what tools should look like in clinical practice.
2) to sensitize consortium partners to the ethical implications of predictive modelling in stroke medicine and have them reflect on their role in this process.
3) to enhance empathy for end-users and beneficiaries of the Precise4Q tools
4) to engage in an in-depth discussion on the potential impact of the Precise4Q tools, with particular focus on identifying the potential ethical challenges and mitigation strategies.

In preparation for the workshop, participants were asked to complete the following two tasks:

> **Task 1:** Please access the collaborative whiteboard and draft a concise (lay) description of what the Precise4Q solution(s) should look like in practice. You may use sticky-notes or insert a text field. When signing in as a visitor, please use your name and/or affiliation.
>
> **Task 2:** Next, please add your notes to the empathy maps. How might the Precise4Q tools we envision impact the reality of patients and clinicians? Creating empathy is a quick way to gain a deeper understanding of who we're designing for.

The second impact workshop series, which was originally planned to coincide with the Precise4Q Plenary meeting in November has been postponed as the Plenary meeting has also been postponed and will now take place in Spring 2022.

## Shared vision of practical tools

Despite a shared common goal and overall vision for Precise4Q i.e., improving health outcomes for stroke survivors and reducing the burden of stroke on the individual and society, it seemed that consortium partners did not share yet a common understanding of what the Precise4Q tool(s) will actually look like in practice. While partners have been refining tools for the individual phases of the stroke patient journey building on earlier work (see also D2.8), it seemed that they have not yet come to a shared understanding of what the tools resulting from Precise4Q will look like and how they will be implemented in practice. From an ethics perspective, this lack of a shared understanding makes it challenging to perform a thorough ethical analysis that is targeted rather than generic. It was therefore a priority for the workshop to advance the consortium's thinking about the more practical application and implementation of the Precise4Q tools.

Before the workshop, participants added sticky notes to the shared vision section of the collaborative whiteboard to describe what the Precise4Q tools should look like in their view. They were asked to particularly reflect on its primary user group, the added value it would bring, the type of

device it would run on, and the way in which information would be presented to end-users. During each of the three workshops, participants then worked in smaller groups to draft a concise (lay) description of what the Precise4Q tools should look like in practice. A template was provided to guide their thinking.

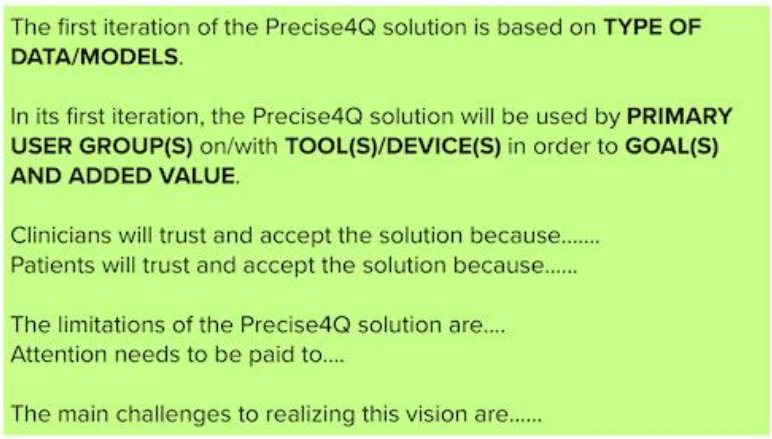


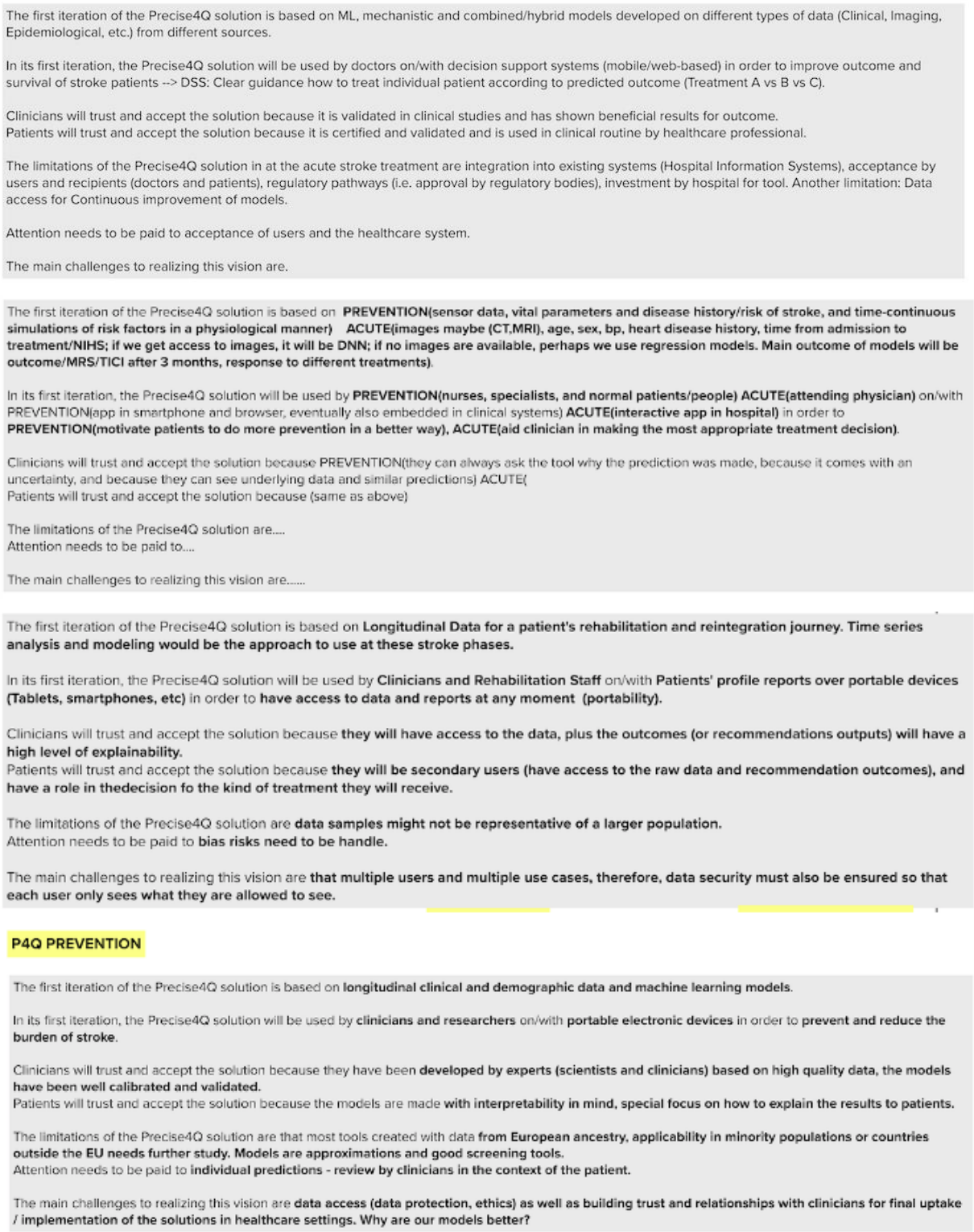

**Figure 2 Template shared vision**

Even though not all groups were able to finalize a detailed description in their small groups, it became evident through the discussion that there was consensus regarding some of the core aspects of the Precise4Q tools (Figure 3).

First, we found that there will likely not be one overarching tool encompassing the entire patient journey but instead individual tools for the different phases of stroke (i.e., prevention, acute, rehabilitation, reintegration) with partly different primary user groups. This finding is also in line with considerations laid out in D1.4 (Set of functional requirements and architecture) and D2.8 (Pilot for clinical decision support system). While most participants identified healthcare professionals as the primary user group, especially in the acute setting, others recognized that patients may also become primary users, especially in the prevention, rehabilitation, and reintegration phase. Also, researchers were considered primary users by some groups, particularly with reference to the European Stroke Research Platform. In terms of devices, we found that partners envision the Precise4Q tools to be installed and used on portable electronic devices (e.g., smartphone, tablet).

The added value was not only seen in facilitating decision-making by providing predictions at the different stages of the stroke patient journey but also in making holistic patient data accessible to decision-makers at the point of care. Participants considered explainability, access to the underlying data, clinical validation, and routinization to be drivers for clinician and patient trust in the Precise4Q tools.

Limitations to the Precise4Q tools were seen in the data sources and related issues of data representativeness (with data being primary from European ancestry). Challenges to realizing the vision were mainly seen in issues related to data access and data availability, integration in hospital information systems (interoperability), regulatory pathways, the need for financial investments from hospitals, ensuring data security, and establishing end-user trust and acceptance.



The first iteration of the Precise4Q solution is based on ML, mechanistic and combined/hybrid models developed on different types of data (Clinical, Imaging, Epidemiological, etc.) from different sources.

In its first iteration, the Precise4Q solution will be used by doctors on/with decision support systems (mobile/web-based) in order to improve outcome and survival of stroke patients --> DSS: Clear guidance how to treat individual patient according to predicted outcome (Treatment A vs B vs C).

Clinicians will trust and accept the solution because it is validated in clinical studies and has shown beneficial results for outcome.
Patients will trust and accept the solution because it is certified and validated and is used in clinical routine by healthcare professional.

The limitations of the Precise4Q solution in at the acute stroke treatment are integration into existing systems (Hospital Information Systems), acceptance by users and recipients (doctors and patients), regulatory pathways (i.e. approval by regulatory bodies), investment by hospital for tool. Another limitation: Data access for Continuous improvement of models.

Attention needs to be paid to acceptance of users and the healthcare system.

The main challenges to realizing this vision are.

The first iteration of the Precise4Q solution is based on **PREVENTION(sensor data, vital parameters and disease history/risk of stroke, and time-continuous simulations of risk factors in a physiological manner) ACUTE(images maybe (CT,MRI), age, sex, bp, heart disease history, time from admission to treatment/NIHS; if we get access to images, it will be DNN; if no images are available, perhaps we use regression models. Main outcome of models will be outcome/MRS/TICI after 3 months, response to different treatments)**.

In its first iteration, the Precise4Q solution will be used by **PREVENTION(nurses, specialists, and normal patients/people) ACUTE(attending physician)** on/with PREVENTION(app in smartphone and browser, eventually also embedded in clinical systems) **ACUTE(interactive app in hospital)** in order to **PREVENTION(motivate patients to do more prevention in a better way), ACUTE(aid clinician in making the most appropriate treatment decision)**.

Clinicians will trust and accept the solution because PREVENTION(they can always ask the tool why the prediction was made, because it comes with an uncertainty, and because they can see underlying data and similar predictions) ACUTE(
Patients will trust and accept the solution because (same as above)

The limitations of the Precise4Q solution are....
Attention needs to be paid to....

The main challenges to realizing this vision are.....

The first iteration of the Precise4Q solution is based on **Longitudinal Data for a patient's rehabilitation and reintegration journey. Time series analysis and modeling would be the approach to use at these stroke phases.**

In its first iteration, the Precise4Q solution will be used by **Clinicians and Rehabilitation Staff** on/with **Patients' profile reports over portable devices (Tablets, smartphones, etc)** in order to **have access to data and reports at any moment (portability).**

Clinicians will trust and accept the solution because **they will have access to the data, plus the outcomes (or recommendations outputs) will have a high level of explainability.**
Patients will trust and accept the solution because **they will be secondary users (have access to the raw data and recommendation outcomes), and have a role in thedecision fo the kind of treatment they will receive.**

The limitations of the Precise4Q solution are **data samples might not be representative of a larger population.**
Attention needs to be paid to **bias risks need to be handle.**

The main challenges to realizing this vision are **that multiple users and multiple use cases, therefore, data security must also be ensured so that each user only sees what they are allowed to see.**

**P4Q PREVENTION**

The first iteration of the Precise4Q solution is based on **longitudinal clinical and demographic data and machine learning models**.

In its first iteration, the Precise4Q solution will be used by **clinicians and researchers** on/with **portable electronic devices** in order to **prevent and reduce the burden of stroke**.

Clinicians will trust and accept the solution because they have been **developed by experts (scientists and clinicians) based on high quality data, the models have been well calibrated and validated.**
Patients will trust and accept the solution because the models are made **with interpretability in mind, special focus on how to explain the results to patients.**

The limitations of the Precise4Q solution are that most tools created with data **from European ancestry, applicability in minority populations or countries outside the EU needs further study. Models are approximations and good screening tools.**
Attention needs to be paid to **individual predictions - review by clinicians in the context of the patient.**

The main challenges to realizing this vision are **data access (data protection, ethics) as well as building trust and relationships with clinicians for final uptake / implementation of the solutions in healthcare settings. Why are our models better?**

**Figure 3 Screenshot: Examples of visions compiled in the smaller working groups**

## Empathy maps

Empathy maps were used to stimulate consortium partners' thinking about the potential implications of the Precise4Q tools and their impact on clinicians, patients, and researchers across the entire patient

journey (i.e., prevention, acute, rehabilitation, reintegration). Moreover, they were a tool used to create empathy for prospective end-users and beneficiaries of the Precise4Q tools and aimed to push consortium partners to reflect on their own role and responsibilities in this process. To this end, workshop participants were asked to reflect on the needs & hopes and fears & concerns of these three stakeholder groups. In addition to working on the empathy maps as a preparatory exercise individually prior to the workshop, empathy maps were also discussed during each of the workshops to identify common themes and discrepancies.

***Patients***

On the patients' side, participants identified the hope and need for personalized care that fits, motivation, and gaining an idea of what the future holds (prognosis), emphasizing the relevance of actionable information tailored to the individual. These considerations reflected a strong aspect of patient empowerment and the opportunities the Precise4Q tool may offer to promote it. There were also references to support being needed from healthcare providers to cope and manage stroke, addressing the hope and need for more relational type of aspects.

Relational issues were also very pronounced in the fears and concerns section. Being just a number, not knowing who is treating, the machine or the human, feeling like a lab rat, a lack of trust in the technology, being alone in need of support were some of the fears and concerns participants suspected patients may have. Their accounts also reflected aspects of patient disempowerment, so loss of control or feeling controlled by an algorithm (the feeling of not being able to keep up with the rehabilitation schedule, for instance), feeling unknowledgeable and overwhelmed by the technology. Participants also recognized that patients may fear that a poor prognosis might lead to a withdrawal or limited access to resources, hinting at aspects of justice and concerns about discrimination. Another theme evolved around privacy concerns, so patients' potential fear of health information becoming public or used for unintended purposes, e.g., by insurance companies or employers.

***Clinicians***

On the clinicians' side, there was a clear focus on the hope to provide the best care possible. Participants suspected that clinicians would hope for the Precise4Q tools to support evidence-based and personalized stroke treatment, by providing better guidelines to help guide decision-making and advancing the general understanding of stroke. There were also aspects related to usability, reflecting participants' considerations that clinicians hope for and need technical tools to be intuitive and easy to integrate into clinical practice. In this context, workshop participants also recognized and emphasized the need for integrating end-users in the development progress as early as possible.

In terms of fears and concerns, participants identified bias to be a major concern to clinicians at all stages. Also, the issue of AI explainability emerged as a central theme, capturing participant's assumptions that clinicians may fears or be concerned about "blackbox algorithms" that provide little information about their inner workings i.e., how predictions or treatment recommendations are derived. Additionally, participants' considered usability and integration in practice to be major concerns for clinicians. An aspect mentioned in the prevention phase was the concern of not knowing how to guide patients through information provided by the Precise4Q tools.

***Researchers***

During the first workshop, it was suggested that researchers themselves are, in fact, also an important user group to consider, which is why a third empathy map was created. Regarding the needs and hopes of researchers for the Precise4Q product(s), workshop participants identified data access and data quality, clinical utility, and adoption in practice is central. Additionally, they considered the need and hope that the Precise4Q product(s) would spur and strengthen collaborations among researchers.

Researchers' fears and concerns were mainly seen in the inability to access data (e.g. due to proprietary or data protection restrictions), as well as data quality and data completeness – aspects which are of direct relevance to their immediate contributions to the Precise4Q tool. Yet, participants also identified concerns about models failing, or simply not being used in practice as potential fears and concerns of fellow researchers. Potential reasons for failing models or limited uptake in practice were seen in models being too expensive or difficult to use (poor usability), or models being not suitable for the setting, for example, because they are too slow in the acute setting. Another concern related to models reinforcing specific practices of a healthcare center that are not applicable to others.

## Challenges & Mitigation strategies

Building on the ethical challenges and questions identified from D1.5 (Empirical study on patients' and clinicians' attitudes towards personalized medicine and multi-dimensional predictive models), workshop participants were asked to identify potential ethical challenges and, in a second step, potential mitigation strategies. Table 1 presents an overview of the ethical challenges, open questions, and mitigation strategies identified by workshop participants.

| Ethical challenges & open questions | Proposed mitigation strategies |
|---|---|
| **Data protection & Privacy** | |
| Personal information can be derived from published models and may pose a threat to patient privacy. What happens to the data and models after the end of the project? If models are not published, who can store them and how long? | Only high-level descriptions of the models or very simplistic models can be published with a special type of procedure to make sure they comply with data protection and privacy regulations |
| **Autonomy** | |
| The Precise4Q solution may pose a threat to clinician autonomy. Especially younger and less experiences clinicians may find themselves under pressure to comply with the system's recommendation. | Decision should always be taken by a (group of) specialized human(s) i.e., clinician(s) |
| Clinicians may blindly trust the system without critically assessing its output against their own knowledge of the case. | Model output needs to indicate the confidence level of the prediction to ensure the clinician doesn't blindly follow or ignores the system's output. |
| Patients need to be in charge of their data and determine who gets access to it (analogy: managing personal finances). | Platforms where patient data is systematically extracted from various sources (EHR, digital twin, etc.) and where patients can then securely store, manage, and share their data as they see fit |

| | |
|---|---|
| What if predictive information is not actionable? (Prevention phase) | Emphasis should be put on modifiable risk factors, i.e., lifestyle factors, like exercising or nutrition, that patients can change. If models get validated enough, they could help find solutions that work for people in their individual contexts. |
| Patient preferences may not be appropriately integrated into the models because often the design of data driven solutions is to a large extent determined by the data that is available to the project | |
| Monopolization of" big data": only big companies have access to data and will thus have the best models – clinicians and patients may have no choice. | |
| **Disclosure** | |
| Full disclosure would be ideal but to fully disclose, clinicians themselves would need a comprehensive understanding and knowledge of the models, also their potential pitfalls and failures. | |
| Should the use of AI be disclosed? And if so, how? How much information should be provided? Can too much information be harmful? | - "Zooming" or "drill-down approach" where patients get simple information first and then continuously more details if they are interested<br>- "Storytelling"<br>- Easy to read material<br>- Never leave a patient alone with their risks, counselling & follow-up is needed<br>- Empathy is needed everywhere but especially in the medical field.<br>- Informed consent (example of vaccination) |
| **Justice** | |
| Less technology-savvy populations may be disadvantaged. | Less technology-savvy people should get more time with clinicians to understand the technology. |
| Predictions could be bias against certain populations. | Introducing feedback loops for continuous improvement of models and error correction. |
| System may replace doctors to reduce costs which may lead to lower income groups not being able to have face to face interactions with clinicians anymore. | |
| **Responsibility** | |
| Uncertainty about who is responsible if patients are harmed. Also, clinicians may put more responsibility on the patient. | - Clarity regarding liability and responsibility is needed.<br>- If models are deployed in clinical care, an entity (e.g., a company) should exist that takes responsibility for adjusting any shortcomings of the models.<br>- The one acting on the system's recommendations, most likely the attending physician, will be responsible. |
| **Non-maleficence** | |
| The impact of predictive information on patient health and well-being is unclear. Poor predictions may lead to additional stress which could cause a stroke instead of helping to prevent it. | Communication of prediction results needs to be further explored |

| | |
|---|---|
| Models may deliver incorrect predictions. | Continuous evaluation might help to catch incorrect predictions (prevention phase) |

**Table 1 Ethical challenges and mitigation strategies**

# 8 Best practices

Through policy monitoring and other initiates in the field of digital ethics and trustworthy AI, we identified three key best practices when it comes to assessing the ethical considerations and impact of medical AI.

**Learning from other use cases**

The assessment of cases or case studies is a commonly used practice in bioethics to pinpoint and address ethical tensions. In AI ethics, we have observed a similar trend i.e., the ethical assessment of concrete examples of applications in development or deployment as an approach to delineate and counteract potential ethical risks of these technologies in practice. While it is important to note that each case is unique and should be regarded in light of its circumstances and boundary conditions, there are some common ethical considerations which may applicable and relevant to future cases as well. Learning for and from practical examples allows researchers and practitioners to move beyond a purely theoretical discourse to an in-depth discussion, enabling them to incorporate learnings from previous assessments thereby increasing rigor and accelerating the assessment process.

**Interdisciplinary collaboration and the integration of ethics in the development process**

Another trend we have observed is the proliferation of interdisciplinary research collaborations to inform the development of ethical AI in healthcare. Researchers coming from various disciplinary backgrounds, including law, medicine, data science, philosophy, and social sciences join forces to combine their subject expertise to inform AI development and assessment The underlying premise here is that breaking down disciplinary silos and fostering interdisciplinary dialogue allows for mutual learning and a more comprehensive view on technological advance without having to reinvent the wheel. Much like Precise4Q includes dedicated work on the ethical, legal, and societal implications of big data health research and medical AI, in an otherwise more technically oriented project, this is the case for an increasing number of international research projects and initiatives in healthcare. In this way, the ethical, legal, and societal considerations can become an integral part of the development process to guide decision making, rather than being a mandatory often tokenistic tag-on at the end of a project. However, interdisciplinary collaboration does not come without challenges as it is characterized by different world views, values, and research priorities. Successful interdisciplinary collaboration therefore requires investment and dedication from all those involved, including the joint establishment of a common reference framework.

**End-user involvement**

Finally, the involvement of end-users and beneficiaries in the development process seems to be a valuable approach to ensuring that technical tools meet the needs of those targeted. In addition to allowing developers to identify potential ethical challenges and risks, involving end-users can also help to increase the trustworthiness, usefulness, and ease of use of technical tools. Involvement can range from one-time-off consultations to more-in-depth involvement, like involving end-users in a co-design process. Particular attention should be paid to involving vulnerable and seldom-heard groups whose views and experiences may differ from the larger part of society.

# 9 Reflective framework

Based on the insights we gathered through the various stages of the project, we conceptualized a draft reflective ethical framework consisting of ten sub-sections across development and deployment. The framework is rooted in the four principles of bioethics: Autonomy, Justice, Beneficence, and Non-maleficence [12] and draws on existing ethical frameworks and guidelines for medical AI [8, 13]. Additionally, it is informed by the lived experience, attitudes, values, and expectations of prospective users, beneficiaries, and developers, as outlined in earlier sections. The reflective framework presented here aims to guide the final stages of development before bringing the PRECISE4Q tools to market and outlines aspects to consider beyond initial deployment (i.e., continuous monitoring and evaluation).

The main purpose of the framework is to stimulate discussion and reflection among the consortium partners, allowing them to identify and anticipate potential ethical challenges that might jeopardize the successful translation of the Precise4Q tools into clinical practice. In doing so, the framework aims to ensure that decision-making during the development and deployment phase is closely aligned with core ethical values and principles of patient-centered care. The framework does not claim to be exhaustive, instead its guiding questions should serve as a basis for interprofessional exchange and reflection within and beyond the consortium, involving also prospective end-users (clinicians) and beneficiaries (patients, family members) to provoke an in-depth engagement which may lead to the identification of new questions to be discussed.

| Phase | Theme | Guiding questions |
|---|---|---|
| Development | Data quality and representativeness | - How data been obtained in an ethical manner?<br>- What measures are in place to ensure data quality and representativeness?<br>- What do we know about the data quality and its representativeness for the target population?<br>- Who might be under or overrepresented?<br>- What consequences may data characteristics have on the performance of the model for these population(s)? |
| | Purpose of the tool | - What is the specific problem the tool aims to address?<br>- What is the intended purpose of the tool in clinical practice?<br>- In what phase of stroke is the tool to be used (prevention, acute, rehabilitation or reintegration)?<br>- How might the tool be used by clinicians and how may this shape their professional role perceptions?<br>- Is there a risk of inappropriate use and how might this risk be mitigated?<br>- Are there any secondary end-user groups? |
| | Explainability | - What kind of information on the tool will be available to end-users?<br>- Are models explainable and if so, is there an impact on performance?<br>- If available, are explanations tailored to the needs of end-users?<br>- How may the information end-users have or lack impact their interaction with the tool? |
| | Usability and user experience | - Have prospective end-users been involved in the development process and if so, how has their input shaped the tool?<br>- If prospective end-users were not involved in the development, what consequences may this have on the tool and its adoption in clinical practice?<br>- Have usability and user experience been assessed, and if so, how? |

| | | |
|---|---|---|
| | Clinical validation | - How is the tool validated?<br>- What does clinical validation mean to developers, what does it mean to clinicians and patients?<br>- What impact may clinical validation have on clinicians' and patients' trust and responsibility? |
| Deployment | Disclosure of AI | - How much information can and should be disclosed to the patient?<br>- How much do clinicians need to know about the tool and its application to fulfil their role?<br>- What impact may predictive health information with disclosure of AI have on patient autonomy, trust, and the doctor-patient relationship (e.g. shared decision-making)?<br>- What about the impact of disclosure on vulnerable populations (e.g., socially disadvantaged groups, stigmatized groups, groups with lower health literacy skills? |
| | Responsibility | - How is responsibility/liability addressed?<br>- Is there a risk of deskilling?<br>- What is the developers' responsibility?<br>- What impact may incorrect decisions caused by the tool have on clinicians' moral responsibility? |
| | Empathy | - How may the tool impact clinicians' empathy towards patients?<br>- How can patient values, beliefs, and preferences be incorporated into the decision-making process?<br>- Might the tool replace human contact in the clinical encounter and if so, what consequences may this have for patients and clinicians? |
| | Privacy & Data Protection | - Given that stroke prevention takes place before any symptoms occur, how can health benefits and privacy be balanced?<br>- Should there be different privacy standards for the different phases of stroke (prevention, acute, rehabilitation, reintegration)?<br>- Which mechanisms would need to be in place to ensure patient privacy?<br>- What might be the consequences of failing to ensure patient privacy? |
| | Monitoring & Evaluation | - What should process and impact monitoring and evaluation look like along the patient journey and life cycle of the technology?<br>- Who is responsible for conducting continuous monitoring and evaluation?<br>- What might be the consequences of failing to conduct continuous monitoring and evaluation? |

# 10 Limitations

The work presented here should be considered in light of some limitations. First, given the rapidly growing body of literature and its fragmentation across disciplines, including computer sciences, medicine, social sciences, and philosophy, the review presented (see 4) may not provide an exhaustive overview of emerging challenges related to multi-dimensional modeling in stroke medicine.

Given the limited scope of qualitative study (see 55), which involved only participants from Germany and Switzerland our findings may not be presentative. In particular, we may have missed capturing the views of vulnerable and disadvantaged groups, including those of immigrants and refugees, for instance. Moreover, our interviews were conducted just prior to the Covid-19 pandemic in late 2019 and early 2020. Considering the tremendous impact, the pandemic had on individuals and healthcare systems worldwide, it may be that it also changed how different stakeholder groups view the role of AI-powered CDSS in stroke medicine and healthcare more generally.

Another limitation can be seen in the fact that the Deliberative Dashboard (see 6) failed to generate the desired engagement. We partly attribute the lack of engagement to the impact the pandemic had on the overall project and consortium partners' prioritization during this time. As attempts to promote uptake were unsuccessful, we launched an internal workshop series to ensure an ongoing interdisciplinary exchange among the consortium partners about the ethical issues related to data-driven predictive modeling in stroke.

Finally, we have to acknowledge that the Precise4Q impact workshop series (see 7) was limited to project-internal participants. This may have biased our findings to exclusively focus on challenges experienced in the specific context of the project, which may not necessarily present a comprehensive picture of developers' and data scientists' views. Also, for the workshop series, it is possible that the extraordinary impact the pandemic had on the project may have shifted priorities and considerations regarding the ethical implications of predictive modelling in stroke, for instance by overemphasizing or neglecting particular aspects.

When conceptualizing the ethical framework (see 8), we tried to account for these shortcomings by drawing on existing ethical frameworks and guidelines for ethical and trustworthy AI in medicine.

# 11 Outlook

The work presented here informs the final deliverable of WP1 (Patients' Needs and Ethical Framework): D1.6 Ethical framework and oversight mechanisms for big data health research. In a next step, we will pilot-test the Precise4Q ethical framework presented here (see 9) and compare it to existing ethical frameworks and checklists for AI development, including the WHO Ethics and governance of Health AI report and the Assessment List for Trustworthy Artificial Intelligence (ALTAI) introduced by the High-Level Expert Group on AI (AI HLEG). In doing so, we aim to explore the consortium partners' experiences with applying the framework and elicit their feedback on how it could be further improved. The pilot-test and evaluation will follow a participatory approach and will be conducted as part of the second Precise4Q impact workshop series (T1.6, see 7) followed by an internal survey (T1.7). Upon refinement and finalization, the Precise4Q ethics framework will be presented at a public symposium (T1.7).

# 12 References


1. Cave, S., et al., *Using AI ethically to tackle covid-19.* bmj, 2021. **372**.
2. Ienca, M. and E. Vayena, *On the responsible use of digital data to tackle the COVID-19 pandemic.* Nature medicine, 2020. **26**(4): p. 463-464.
3. Vinay, R., H. Baumann, and N. Biller-Andorno, *Ethics of ICU triage during COVID-19.* British Medical Bulletin, 2021. **138**(1): p. 5-15.
4. Hantrais, L., et al., *Covid-19 and the digital revolution.* Contemporary Social Science, 2021. **16**(2): p. 256-270.
5. Blasimme, A. and E. Vayena, *What's next for COVID-19 apps? Governance and oversight.* Science, 2020. **370**(6518): p. 760-762.
6. Allahabadi, H., et al., *Assessing Trustworthy AI in times of COVID-19. Deep Learning for predicting a multi-regional score conveying the degree of lung compromise in COVID-19 patients.* IEEE Transactions on Technology and Society, under review.
7. Epstein, R.M. and R.L. Street, *The values and value of patient-centered care*. 2011, Annals Family Med. p. 100-103.
8. HLEG. *High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI.* . 2019 [cited 2021 10.12.2021]; Available from: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.
9. EC. *ALTAI - The Assessment List on Trustworthy Artificial Intelligence*. 2020; Available from: https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence.
10. Zicari, R.V., et al., *Co-Design of a Trustworthy AI System in Healthcare: Deep Learning Based Skin Lesion Classifier.* Frontiers in Human Dynamics, 2021: p. 40.
11. Zicari, R.V., et al., *On Assessing Trustworthy AI in Healthcare: Machine Learning as a Supportive Tool to Recognize Cardiac Arrest in Emergency Calls.* Frontiers in Human Dynamics, 2021. **3**: p. 673104.
12. Beauchamp, T.L. and J.F. Childress, *Principles of biomedical ethics*. 2001: Oxford University Press, USA.
13. Floridi, L., et al., *AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations.* Minds and Machines, 2018. **28**(4): p. 689-707.