LSE Consulting

Evaluation framework for health professionals' digital health and AI technologies

Evidence-based policy recommendations

Robin van Kessel, Jelena Schmidt, Stephanie Winitsky, George Wharton, Elias Mossialos

October 2025

LSE Consulting

+44 (0)20 7107 5215
consulting@lse.ac.uk
lse.ac.uk/consultancy

Please cite this study as: Robin van Kessel, Jelena Schmidt, Stephanie Winitsky, George Wharton, Elias Mossialos, 2025. 'Evaluation framework for health professionals' digital health and AI technologies: Evidence-based policy recommendations. DOI: 10.21953/lse.vi7ayokh6s1u

This report is commissioned via LSE Consulting which is set up by The London School of Economics and Political Science to enable and facilitate the application of its academic expertise and intellectual resources.

LSE Enterprise Ltd, trading as LSE Consulting, is a wholly-owned subsidiary of The London School of Economics and Political Science. The LSE trademark is used under licence from The London School of Economics and Political Science.

LSE Consulting
Houghton Street
London WC2A 2AE

Acknowledgements

This research has received financial support from Roche Diagnostics. The authors are grateful to Afua van Haasteren, Katalin Ersek, Paola Botrugno, Iren Yaprak Yay, Chaohui Guo from Roche Diagnostics for their helpful input and support during the development of this report.

Disclaimer

The field of digital health and artificial intelligence is rapidly advancing. Sources published after February 14th, 2025 were not used in the development of this report.

About the authors

Robin van Kessel, PhD is the Hoffmann Fellow between LSE Health, London School of Economics (LSE) and the World Economic Forum and the founder of LSE Health Digital.

Jelena Schmidt, MSc is a Health Policy Research Associate at LSE Health Digital, London School of Economics (LSE).

Stephanie Winitsky, MSc is a Health Policy Associate at LSE Health, London School of Economics (LSE).

George Wharton, MSc is an Associate Professor (Education) and Programme Director in the Department of Health Policy, London School of Economics (LSE).

Elias Mossialos, PhD is Cheng Yu Tung Chair in Global Health and Director of LSE Health, London School of Economics (LSE).

Table of Contents

List of Tables

List of Figures

List of Boxes

Abbreviations

NCDs	Non-communicable diseases
DHAITs	Digital health and artificial intelligence technologies
HCP	Healthcare professional
EU	European Union
IVD	In-vitro diagnostic
IMDRF	International Medical Device Regulators Forum
USA	United States of America
SaMD	Software as a Medical Device
UK	United Kingdom
NICE	National Institute for Health and Care Excellence
HAS	Haute Autorité de Santé
FDA	Food and Drugs Administration
CDS	Clinical decision support
RCT	Randomized controlled trial
DiGA	Digitale Gesundheitsanwendungen
QALY	Quality-adjusted life-years
CMA	Cost-minimalization analysis
CCA	Cost-consequence analysis
HTA	Health technology assessment
PCCP	Predetermined Change Control Plans
MDR	Medical Devices Regulation
IVDR	In-vitro Diagnostic Devices Regulation
EHR	Electronic health record
MHRA	Medicines and Healthcare products Regulatory Agency
CE	Conformité Européenne

Executive Summary

Healthcare systems around the world are facing mounting financial pressure and are increasingly being stretched across competing priorities. At the same time, the global shortage of healthcare workers is accelerating. These challenges are compounded by the persistent rise in noncommunicable diseases (NCDs), which are now responsible for 41 million deaths annually – 74% of all global mortality. Digital health and artificial intelligence technologies (DHAITs) hold considerable promise in addressing these pressures. By enabling more accessible, sustainable, efficient and higher quality care. While attention often centers on patient-facing tools, digital solutions used by healthcare professionals (HCPs)—including clinicians, nurses, managers, and administrators—are equally important. These tools support critical functions such as risk analysis, screening, diagnosis and prognosis, treatment choices, and patient monitoring, with potential to optimise workflows, reduce unwarranted variation in care, and improve both provider efficiency and patient outcomes.

Despite their potential, DHAITs continue to face systemic adoption barriers. Many healthcare providers and patients lack the necessary digital literacy to effectively use these tools, and infrastructure constraints and resource limitations are widespread. Even though the implementation of the European Health Data Space aims to create a common framework for the use and exchange of health data across the European Union (EU), the current absence of standardised frameworks for data access, sharing, and governance contributes to fragmentation across systems globally. The lack of interoperability between different systems complicates their integration and discourages healthcare professionals to adopt DHAITs. Most significantly, the sector lacks robust, context-sensitive evidence demonstrating long-term value. Existing evaluation models, inherited from the pharmaceutical and medical technology (including medical devices and IVD diagnostics) sector, often prove ill-suited for digital tools, which tend to be iterative, adaptive, and faster-moving. As a result, many digital solutions fail to achieve scale, leading to a proliferation of short-lived, low-value tools that never realise their intended impact.

Objective

This report sets out to strengthen the foundation for evaluating healthcare professional-facing DHAITs. It introduces an evidence-based taxonomy designed specifically for tools used by healthcare professionals. Additionally, it maps current evidence frameworks across a selection of different health systems to support the development of a future-fit assessment methodology.

Classification of digital health and AI technologies

Most regulatory authorities align with the International Medical Device Regulators Forum (IMDRF) framework for classifying medical devices, which classifies medical devices based on their risk level and clinical purpose. While broadly aligned, variations remain—for example, the United States of America (USA) recognizes three risk classes (I–III), whereas the EU subdivides class II into IIa and IIb.

DHAITs are widely being classified as a Software as a Medical Device (SaMD), though it does not capture the full spectrum of DHAITs. To address this, several jurisdictions are advancing purpose-based or function-based classification frameworks:

- United Kingdom (UK): The National Institute for Health and Care Excellence (NICE) applies a three-tier framework distinguishing system-, management-, and intervention-focused technologies, aligning evidence requirements with potential user/system risk.
- France: The *Haute Autorité de Santé* (HAS) adds a fourth tier for technologies capable of autonomous decision-making with limited human oversight.
- South Korea: Classification depends on intended use and risk of harm, particularly how software influences clinical decision-making.
- USA: The Food and Drugs Administration (FDA) treats clinical decision support (CDS) tools as non-devices if they support but do not drive decisions and allow independent clinician review.
- Canada: Health Canada adds functional distinctions between medical-grade AI and consumer health tools.

Taxonomy for professional-facing digital health and AI technology

This report introduces a taxonomy based on building blocks developed through scoping reviews and thematic analysis to provide a structured yet flexible classification system. It is not meant to be final but provides a common framework to describe core features shared by many HCP-facing technologies.

The taxonomy focuses on seven core dimensions in descending order from specific context to general foundation:

1) Interoperability: ability of the DHAIT to interact with other digital and data systems either as a standalone software or being embedded in existing digital infrastructures.
2) Access platform: access to the DHAIT is possible via mobile phone, browser, immersive technologies (e.g., virtual or augmented reality), or hospital IT software.

3) Driving technology: rule-based software or AI-driven models, including machine learning algorithms.
4) Data inputs: real-world or research data.
5) Intended impact: with a distinction between whether the benefits of the DHAIT target the professional, the patient, and/or the health system.
6) Intended use case: including diagnosis, management, monitoring, treatment, prognosis, and prevention.
7) Intended beneficiary: healthcare professionals, including physicians, general practitioners, nurses, pathologists, and allied health workers.

Evidence Requirements for professional-facing digital health and AI technologies

Evidence standards for DHAITs vary by risk level and intended use. Low-risk, non-interventional tools typically require only usability and performance testing. In contrast, AI tools used for diagnostics or clinical decision support must demonstrate comparative effectiveness. High-risk systems, including autonomous AI, face the most stringent requirements: prospective trials, real-world evidence, and regulatory sandboxing.

Randomized controlled trials (RCTs) remain the gold standard across the six countries for evaluating clinical evidence. However, given the rapid iteration of digital technologies, alternatives are increasingly accepted:

- UK: NICE allows pragmatic and observational studies when RCTs are unfeasible.
- France: HAS supports alternative designs such as randomized consent trials and target trial emulation.
- Germany: Under the *Digitale Gesundheitsanwendungen* (DiGA) fast track process, real-world evidence is accepted in place of RCTs. Though DiGA targets patient-facing apps, its evidence standards may inform approaches for HCP-facing technologies.

Cost-effectiveness analysis using quality-adjusted life-years (QALYs) remains the dominant choice for assessing economic evidence. However, this method is less suited to HCP-facing tools, as these often yield indirect benefits, and countries have started to propose alternative methods:

- UK & France: Cost-minimisation analysis (CMA) is accepted when clinical equivalence is demonstrated.
- UK & France: Cost-consequence analysis (CCA) is conditionally accepted for tools offering indirect system benefits, such as workflow efficiency.

Health technology assessment (HTA) frameworks have historically prioritized patient outcomes, which may limit their applicability to HCP-facing digital tools. The DiGA framework offers a potential path forward by acknowledging indirect patient-relevant outcomes, including procedural improvements and system-level efficiencies.

Evidence thresholds and uncertainty

Some AI tools are approved in the United States and South Korea at a regulatory level based on substantial equivalence to previously authorised technologies. Although the substantial equivalence approach has not been formally adopted in Europe, there are emerging provisions for reducing the trial burden where prior evidence exists.

To manage the continuous evolution of these technologies, the FDA's Predetermined Change Control Plans (PCCPs) enable manufacturers to pre-authorise updates to AI algorithms. Canada and South Korea allow conditional updates and rely on post-market surveillance to ensure ongoing safety and effectiveness. The EU is in its early stages of policy development with the EU AI Act, which introduces a similar provision to the FDAs PCCPs as well as post-market monitoring for high-risk AI systems.

Uncertainty in health economic models remains a common concern across jurisdictions. Sensitivity analyses are universally required, while probabilistic models are standard in the UK and France. French authorities have also begun incorporating Bayesian modelling, particularly in cases where RCTs are impractical.

Policy recommendations

1) Develop functional classification frameworks for HCP-facing DHAITs aligned with risk profiles.
2) Systematically link classification categories to tailored evidence standards and fit-for-purpose metrics.
3) Prioritise Bayesian statistical methods for the evaluation DHAITs.
4) Implement PCCPs for AI-based technologies.
5) Align HTA, regulatory, and resource allocation processes through early stakeholder engagement and shared data standards.
6) Broaden relevance HTA processes to non-patient-facing DHAITs by incorporating outcome metrics capturing system-level and professional-facing impacts, such as efficiency gains and workflow optimization.
7) Pilot and refine frameworks through multistakeholder engagement.
8) Leverage international regulatory networks for regulating and evaluating healthcare professional facing DHAITs.

Introduction

Health systems are under financial pressure, with competing priorities squeezing the public funds available for health . While the population is ageing, the shortage of healthcare workers is increasing, which, at the current pace, will have an estimated gap of 10 million personnel globally by 2030. Adding to worsening shortages of healthcare workers, is the inequality in geographical distribution of the workforce between and within countries and regions, particularly between urban and rural areas [1,2]. This geographical disparity between countries can be seen in the field of pathology, where two-thirds of all pathologists are located in just ten countries [3]. At the same time, the burden of non-communicable diseases continues to increase, killing 41 million people every year, representing 74% of all deaths and most premature mortality worldwide, and straining the healthcare system [4].

Digital health and artificial intelligence technologies (DHAITs) offer significant potential to address major healthcare system challenges by improving accessibility, sustainability, efficiency, and quality [2,5]. They are widely recognised as foundational for economic growth and cross-cutting accelerators for achieving the Sustainable Development Goals [6]. Investing today an additional US$0.24 per patient per year in digital health interventions, such as telemedicine, mobile messaging and chatbots, is said to help save more than 2 million lives from non-communicable diseases over the next decade and lead to an additional US$ 199 billion in economic benefits. The aversion of approximately 7 million acute events and hospitalizations could reduce the strain on healthcare systems worldwide [7]. DHAITs are not solely used by patients but also by their relatives, healthcare providers, including pathologists, health management personnel, or data services [8,9].

The adoption of DHAITs in healthcare systems is limited by a lack of skills and expertise among end-users (i.e., patients, professionals), resource constraints, emerging digital infrastructure, and the absence of common standards for data management (e.g., siloed data, concerns about health data access and sharing) [10,11]. The lack of interoperability between different systems, for example between those used in primary and secondary care, complicates their integration and discourages healthcare professionals to adopt DHAITs. Within the European Union (EU), the challenge of data standards is being targeted by the implementation of the European Health Data Space, which aims to create a common framework for the safe and secure use and exchange of health data across the EU [12]. Furthermore, strong evidence regarding the real benefits and impacts on health systems and individual well-being remains scarce [13], leading to the development and introduction of low-value, short-lived DHAITs [14].

Due to the novelty and the fast-changing landscape of DHAITs, identifying suitable methods for regulating and evaluating these tools is an ongoing process [15]. Evaluation methods from the field of pharmaceuticals and medical technology (including medical devices and

IVD diagnostics) may serve as a base for DHAIT-specific evaluation frameworks. Still, they cannot be completely transferred to DHAITs [16] because of their continuous/lifetime learning capabilities [13], their rapid technical innovation, and lack of adequate comparators [17]. Similar to the evaluation methods for pharmaceuticals, existent DHAIT-specific HTA focuses predominantly on health-related benefits and cost-utility analyses [18] and do not translate well to non-patient-facing technologies. Therefore, it becomes crucial to develop value assessment frameworks that align with the specific nature of HCP-facing DHAITs.

There are different initiatives ongoing, such as AssessDHT and EDiHTA, which are focusing on developing an harmonized evaluation framework for DHTAIs in the European Union and which are including a more holistic perspective [19,20]. However, with the ongoing development of novel DHAITs and their broad application, it is important to clarify the breadth and depth of HCP-facing DHAITs before developing evaluation frameworks. While some previous work on digital health taxonomies included healthcare professionals facing technologies [8,21,22], these taxonomies differentiate based on the intended beneficiary (most of the time, the patient) and not the intended user. Although healthcare providers are included as target groups, these frameworks are not specifically tailored to the nuances of professional-facing DHAITs [21]. The lack of standardized definitions regarding their functions and user roles significantly hinders not only their use [23] but also the development of tailored assessment criteria.

Objective

The aim of this report is two-fold. First, an evidence-based taxonomy for healthcare professionals facing DHAITs has been developed to enable the development of more targeted and appropriate evaluation methods. Secondly, to inform the evidence assessment framework for HCPs-facing DHAIT, a policy analysis was conducted to assess the current evidence requirements for medical devices and technologies as a whole and the requirements explicitly developed for DHAITs. The policy analysis focussed on evidence requirements in six countries (Canada, France, Germany, South Korea, United Kingdom, United States), that were purposely sampled to represent a mix of different political and health systems, as well as regions, country sizes and levels of digital maturity. The policy analysis of DHAITs was not limited to HCP-facing technologies as DHAIT-focused frameworks are still nascent, and even evidence guidance for patient- or system-facing DHAITs can give important insights into national approaches to establishing evidence requirements for HCP-focused DHAITs.

Classification of digital health and AI technologies

All six examined countries draw upon the risk classification framework for medical devices of the International Medical Device Regulators Forum (IMDRF), which ensures that oversight is proportionate to potential harm. This global framework also covers Software as a Medical Device (SaMD). Based on this classification framework, a medical device can be used for a myriad of purposes, including but not limited to diagnosing, predicting, monitoring, or treating diseases using clinical data or providing critical clinical information. Even though the six studied countries are broadly aligned with the IMDRF risk-based framework, some discrepancies exist. For instance, the risk classification framework in the USA recognizes three risk classes (class I, low risk; class II, moderate risk; class III, high risk), whereas the framework in the EU further split class II, moderate risk, into class IIa, low-to-moderate risk and class IIb, moderate-to-high risk.

DHAITs are currently often described under the umbrella SaMDs, though this single umbrella term hardly captures the heterogeneity and diversity of these technologies, ranging from wearable sensors to AI-enabled digital diagnostic tools. To better delineate between different forms and types of DHAITs, several policy approaches to developing function- or purpose-based classification frameworks have been observed:[1]

- The National Institute for Health and Care Excellence (NICE) in the United Kingdom built a three-layer framework distinguishing between system-, wellbeing-, and care-focused DHAITs. According to NICE, classifying DHAITs by intended purpose allows them to be pooled into risk categories based on the potential risk to service users and the system. The evidence level needed for each tier can then be proportionate to the potential risk to service users from the DHAITs in that tier.
- The *Haute Autorité de Santé* (HAS) in France developed a similar framework but added a fourth layer specifically aimed at DHAITs capable of autonomous decision-making with limited human intervention.
- The Medical Device Evaluation Department of the Ministry of Food and Drug Safety in South Korea specifies in a guidance document that the classification of software using machine learning as a medical device is determined by its intended use and potential hazard risks (i.e., whether the software can cause harm to a patient and its influence on the decision-making of the healthcare professional).

[1] We acknowledge that the German DiGA framework is not listed here, as it does not set out how digital health technologies should be classified from a risk or purpose perspective.

- The Food and Drugs Administration (FDA) in the United States of America delineates clinical-decision support (CDS) systems as a regulatory boundary rather than embedding it within a classification framework. The US FDA stipulates that CDS systems do not qualify as medical devices if they enable clinicians to assess recommendations independently, do not drive clinical decisions, and are intended for lower-risk applications.
- Health Canada introduces additional functional considerations in its evaluation of SaMD, drawing a regulatory distinction between medical-grade AI and consumer health applications.

Until now, DHAITs have been granted market authorisation in Europe and the United States of America by being considered SaMDs. The vast majority of these technologies were labelled as class IIa under the European Union risk classification and low-risk under the risk classification of the United States of America [24,25]. With the introduction of the EU AI Act, AI systems in regulated digital medical products, such as those in AI/ML-enabled medical devices, are classified as high-risk in the EU [15,26,27]. Consequently, the market access pathway of DHAITs will be regulated by the Medical Devices Regulation (MDR), In-Vitro Diagnostic Devices Regulation (IVDR), and AI Act simultaneously. The monitoring of their interaction and alignment is an ongoing process. The emergence of purpose-based classification frameworks to help disentangle the umbrella of SaMDs into more specific categories helps prioritize and manage the review of health technologies. It considers the potential risks and intended use of a technology to determine the level of scrutiny needed during evidence assessment.

Evidence-based taxonomy for professional-facing digital health and AI technologies

There is no standardized definition of what constitutes a healthcare professional facing DHAIT. To bring clarity to this space, a taxonomy has been developed to provide a structured yet adaptable way of understanding DHAITs used by healthcare professionals (see Figure 1). The taxonomy development is based on a scoping review and was guided by a thematic analysis. This iterative and flexible process enabled the identification of recurring concepts and themes across the collected data.

It has to be noted that this taxonomy is not intended to be definitive or exhaustive. Instead, it offers a shared language to describe key characteristics that many HCPs-facing DHAITs may have in common. It is built around seven dimensions: intended beneficiary, intended use case, intended impact, data inputs, driving technology, access platform, and interoperability. Due to the multitude of different types of digital health or AI technology, the taxonomy does not use arrows but different building blocks, which can be expanded, combined, interchanged, and removed as our understanding of these technologies grows. While each DHAIT is unique, this framework can help guide analysis, facilitate comparisons, and support strategic decision-making.

Figure 1: Taxonomy of health professional-facing digital health and AI technologies.

The base of the taxonomy focuses on DHAITs that are used directly by healthcare professionals. The "intended beneficiary" may include physicians, general practitioners, nurses, medical assistants, pathologists, and other clinical staff involved in patient care and decision-making.

DHAITs may be designed to support healthcare professionals in a range of functions. The "intended use case", which is based on the EU MDR/IVDR classification for class IIa devices [28], encompasses the following six functions:

1) *Diagnosis*: digital health or AI technology assists in identifying diseases or medical conditions [29][28];
2) *Management*: digital health or AI technology supports healthcare professionals in organizing and optimizing care [30];
3) *Monitoring*: the digital health or AI technology enables continuous tracking of patients health [31];
4) *Treatment*: digital health or AI technology assists the healthcare professional in delivering therapeutic interventions [32];
5) *Prognosis*: digital health or AI technology helps predict the course of a disease or patient outcome [33]. lastly,
6) *Prevention*: digital health or AI technology assists in reducing the likelihood that a disease or disorder will affect an individual by identifying high-risk individuals [34].

Complementing the intended use case is the “intended impact” category, which defines the expected results the DHAITs aims to achieve. These two categories are connected as the impact of the DHTAIs depends on how and where it is being used. The intended impact can be separated into professional, patient, and system levels. On the healthcare professional level, the technology can aim to streamline disease monitoring [35], build relevant skills, for example to treat more specialized diseases [36], or help deliver higher-quality treatment [37]. On the patient level, it may impact patients’ health literacy skills, patient satisfaction [38], ability to self-manage their condition adherence to the treatment. Furthermore, it supports the detection and diagnosis of diseases, possibly leading to better health outcomes in the long term. At a system level, DHAITs can support efficiency [39], allocate resources better, contain costs [40], and increase coverage [30].

The DHAITs are inherently data-driven, requiring different “data inputs” for processing and analysis. These data inputs can be real-world data collected in the healthcare system, including Electronic Health Record (EHR) data, which consists of patient information, such as medical history, prescriptions, and clinical notes [41]; medical images, which provide visual data from radiology, pathology, and other imaging modalities [29], and administrative data. Real-world data can also come from laboratory information systems or directly from the patients. Patient-reported data refers to self-reported information from individuals regarding their symptoms, experiences, and treatment responses [42]. Other data input can be research data coming from clinical trials or surveys.

The “driving technology” category defines the fundamental technical approaches that power digital health solutions. Some technologies are based on deterministic software. Others may rely on machine learning and artificial intelligence (AI), enabling dynamic, intelligent decision-making through advanced computational techniques. These AI models differ based on the required input and the output they produce, which underlines the importance of viewing AI-based technologies within the context of their purpose and not construing AI as a monolith. Supervised algorithms rely on labelled datasets for training and learning by mapping inputs to known outputs, making them effective for tasks like disease classification. Artificial neural networks, modelled after the human brain, excel at complex, non-linear pattern recognition. Convolutional neural networks are particularly effective for processing images and spatial data [43], whereas recurrent neural networks specialize in sequential data analysis, making them suitable for tasks like time-series forecasting [44]. Transformer models have shown strong performance in handling unstructured data and are used to generate medical documentation from voice input [45].

The “access platform” category refers to the medium through which DHAITs are deployed and accessed by the healthcare professional. Mobile phones can serve as a basis for applications and tools designed for smartphones and tablets to enable point-of-care

diagnostics, like a smartphone-assisted direct ophthalmoscope camera [46]. Web browsers can access electronic health record systems or AI-powered analytics [41]. Furthermore, healthcare professionals can use wearable devices, such as patch heart rhythm monitors, to support the identification of health abnormalities [47]. Additionally, virtual or augmented reality offers immersive technologies used for training [48] or treatment support [49].

Finally, the "interoperability" category encapsulates the extent to which a DHAIT is able to function within and communicate with other parts of the healthcare ecosystem – either as a standalone technology or as a technology being embedded within other parts of the digital infrastructure (e.g., EHR systems). We delineated five practical levels of interoperability:

1) *View-Only Access*: displaying EHR data via a web portal or static reports within the tool, often in a unidirectional way.
2) *Discrete Data Retrieval*: the ability to select and retrieve clearly defined data from another part of the digital infrastructure.
3) *Bidirectional Data Exchange*: the ability to both import and export relevant data to other parts of the digital infrastructure.
4) *Workflow Integration*: embedding a DHAIT within existing infrastructures, potentially launching it from the EHR, utilizing EHR user and patient context, and triggering alerts or notifications within the EHR environment.
5) *Semantic Interoperability*: data exchanged is understood at a conceptual level by both systems, enabling advanced functions like real-time clinical decision support based on dynamic EHR data, AI-driven insights integrated into care pathways, or automated analytics drawing from EHRs.

An example of how this taxonomy could be applied in practice is shown in Box 1 [50].

Box 1: Exemplary assemble of the evidence-based taxonomy.

Deep learning model: 3-dimensional (3-D) Convolutional Neural Network called HeadXNet for segmentation of intracranial aneurysms from CT scans

Interoperability: Not reported

Access platform: Browser

Driving Technology: Machine Learning and Artificial Intelligence – convolutional neural network

Data Inputs: Real world data – medical images

Intended impact:

- Professional level: disease monitoring, reatment quality
- Patient level: disease detection and diagnosis
- System level: efficiency gains

Intended use case: Diagnosis – to automatically detect intracranial aneurysms on CTA and produce segmentations specifying regions of interest

Intended beneficiary: Healthcare professional

Evidence Requirements for professional-facing digital health and AI

Research designs and endpoints for clinical and economic evaluations

Low-risk and non-interventional technologies, such as well-being and self-management applications, typically require minimal clinical validation, often relying on usability and performance testing rather than formal clinical trials. AI-driven diagnostics and decision-support tools, particularly those intended to influence clinical decision-making, are subject to comparative effectiveness studies that assess performance against existing standards of care. Technologies classified at the highest functional level, such as autonomous AI systems or high-risk interventions, face the most stringent evidence requirements, including prospective clinical trials, real-world data collection, and regulatory sandbox evaluations before widespread adoption.

In terms of evidence requirements (see Figure 2), randomized controlled trials (RCTs) remain the gold standard research design across all six countries, with national HTA agencies consistently emphasizing their importance in assessing the safety and effectiveness of AI-driven health interventions. However, given the challenges of conducting RCTs for software-based technologies that evolve rapidly through iterative updates, several jurisdictions have introduced alternative study designs to ensure timely yet rigorous evaluation (see Table 1):

- In the UK, NICE updated its evidence standards framework in 2022 to acknowledge that traditional RCTs may not always be feasible or appropriate for DHAITs and to support the use of alternative study designs, such as pragmatic trials and observational studies, to generate relevant evidence [51].
- In France, the HAS has explicitly allowed alternative methodologies when an RCT is impractical. The HAS permits randomized consent trials, pragmatic trials, and target trial emulation. These approaches have been increasingly accepted for DHAITs, particularly in cases where real-world validation is more appropriate than traditional trial structures [52,53].
- Germany, under its DiGA fast-track process, allows the use of real-world evidence in place of RCTs, enabling continuous assessment through dynamic HTA processes [16]. Under the DiGA framework, digital health applications can be categorised as low- or higher-risk medical devices used to support the detection, monitoring, treatment, or alleviation of diseases or the detection, treatment, alleviation, or compensation of injuries or disabilities. It has to be noted that the DiGA framework is designed for patient-facing digital health applications. Even so, it is important to highlight this framework as it

provides initial evidence standards for DHAITs that can become relevant for HCP-facing technologies.

Figure 2: Evidence provisions for digital health and AI technologies in health technology assessment guidelines.

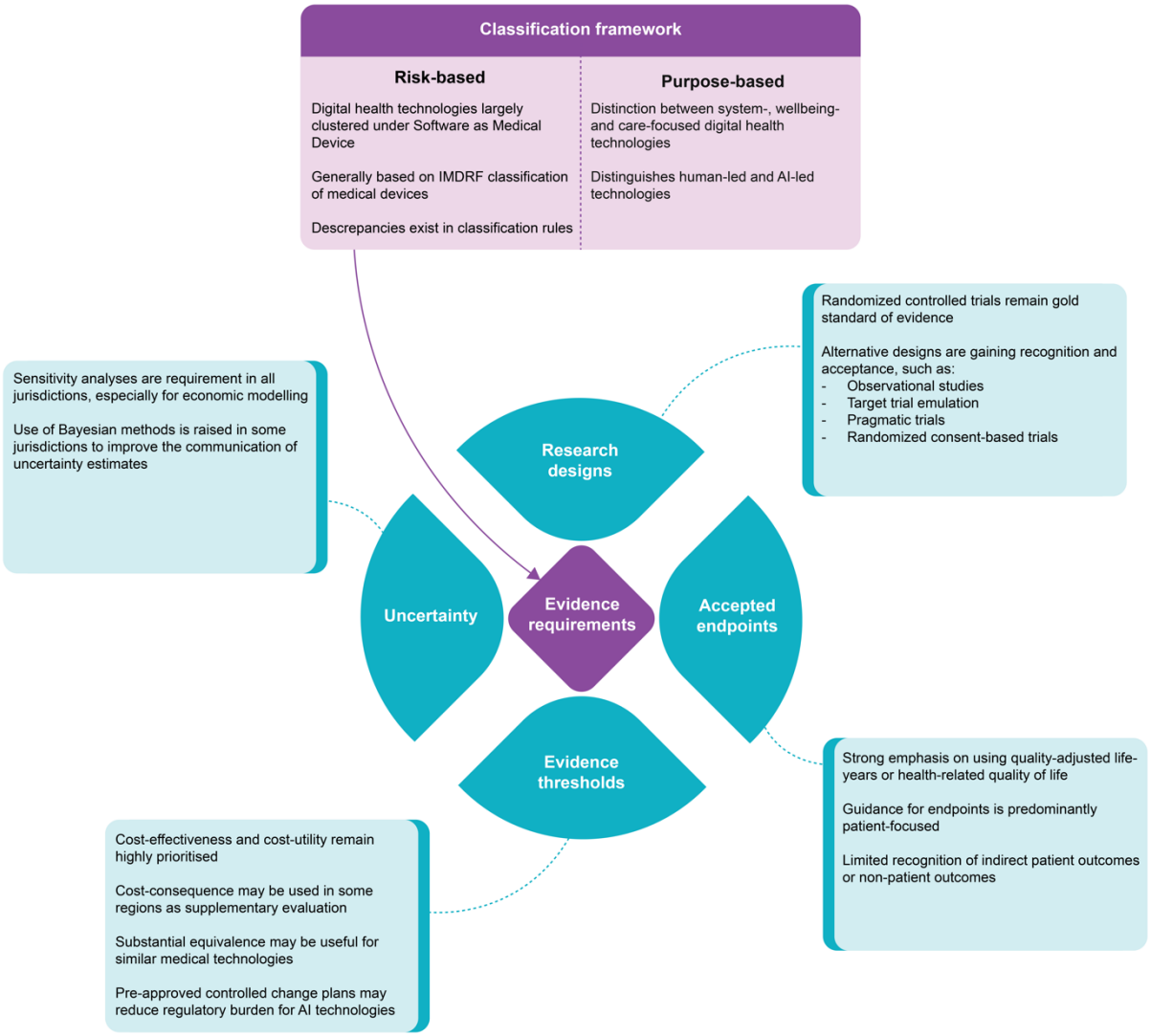

Table 1: Overview of the alternative study designs and their application to digital health and AI technologies.

Study Design	Deviation from traditional RCT design	Benefits for digital health and AI technologies	Example in digital health and AI technologies
Clinical Simulation [54]	Within clinical simulation research the study population is being situated in realistic clinical scenarios to perform tasks close to real-life environments. It does not use randomisation and focuses on testing and evaluating a solution.	Flexible, adaptable, scalable, and relatively inexpensive approach. Possibly useful during the initial stages of the DHAIT development process, which can then cascade into some form of clinical trial.	The NAVIFY Tumor Board solution was tested within a series of simulated lung cancer multidisciplinary team (MDT) meetings for its usability, quality, required time, and the cognitive burden on participants while using the tool. In the simulated MTD sessions the participants discussed up to 10 synthetic patient cases. First without using standard tools, and second with the NAVIFY Tumor Board solution. [54]
Cluster randomized study [55]	Cluster RCTs randomize groups (e.g. clinics, hospitals, communities) instead of individual participants to different interventions arms.	Suited for digital solutions supporting group efforts. Useful when the intervention consists of multiple individual elements, which are influencing or interacting with each other, as direct and indirect effects of an intervention can be evaluated.	DYNAMIC study: Cluster randomized, open-label trial to evaluate the use of a digital clinical decision support algorithm (CDSA), enhanced by point-of-care tests, training and mentorship, compared with usual care, among sick children 2 to 59 months old presenting to primary care facilities for an acute illness in Tanzania [56].
In-silico clinical trial [57]	In-silico clinical trials use virtual patients and computer simulation to predict intervention effects.	Flexible, adaptable, scalable, and relatively inexpensive approach. Useful during the initial stages of the DHAIT	Not yet used.

		development process, which can then cascade into clinical trials.	
Micro-randomized study [55]	Micro-randomized studies randomize interventions multiple times per participant (e.g., daily or hourly) to evaluate the short-term effects of just-in-time adaptive interventions.	Suited to determine the efficacy of a specific component of an DHAITs. Usable especially in the early stages of a product development.	To optimize HeartSteps, a MRT was conducted to assess the efficacy of HeartSteps' contextually tailored activity suggestions. The different activity suggestions were micro-randomized for each participant at each of the five decision points on each day of the study [58].
Pragmatic Trial [59]	A pragmatic trial is an RCT conducted under typical real-world conditions rather than tightly controlled settings. Pragmatic trials use broad inclusion criteria and embed the intervention into routine care with minimal extra monitoring or protocol-mandated restrictions. Blinding may be relaxed, and multiple real-world sites may be involved.	Yields high external validity and real-world clinical effectiveness data. Demonstrates how an AI tool or app performs in diverse patient populations and workflows. Supports cost-effectiveness through the use of existing infrastructure and representative outcomes.	*BlueStar* diabetes app trial: This multicenter pragmatic RCT tested a mobile health intervention in routine practice. While it found no significant overall reduction in Hba1c, it illustrated the real-world performance of a digital therapeutic [46].
Pre-post implementation study [55]	Pre-post implementation studies compare outcomes before and after an intervention in the same group, without a control group.	Suitable to test short-term outcomes.	A computerised physician order entry of chemotherapy order (C-CO) with clinical decision support system was compared six months after its implementation to paper-based chemotherapy order (P-CO) in its usability, feasibility, and efficiency [60].
Observational Trials [59]	No randomization; researchers observe	Provides real-world evidence of	*Naluri* digital health coaching: A

	outcomes as they occur in practice. This includes prospective cohorts, registries, or retrospective EHR analyses. Higher risk of confounding compared to RCTs.	effectiveness and cost-effectiveness at scale. Useful for assessing digital health tools using EHR or claims data. Reflects actual use and outcomes in broad populations.	retrospective observational study showed significant health improvements among app users compared to non-users in a workplace setting [61].
Target Trial Emulation [62]	Emulates a hypothetical RCT using observational data. Defines eligibility, interventions, outcomes, and follow-up as if conducting an RCT, but without randomization. Reduces bias through protocol-like structure and confounding adjustment.	Enhances credibility of real-world evidence. Allows assessment of AI/digital tools post-deployment using large-scale datasets. Faster and cheaper than prospective trials while approximating causal inference.	Not yet available.
Zelen's Design [63]	Randomizes patients before consent. In single-consent version, only intervention group is informed and consents. The control group gets standard care, often without knowing they were in a study. Intention-to-treat analysis applied.	Simplifies recruitment and reflects routine care offer of digital tools. Minimizes disappointment and contamination bias. Suitable for low-risk interventions with variable adherence. May support generalizability and policy relevance.	*MotivATE* web intervention [64]: A Zelen RCT invited only intervention arm participants to access the tool. The trial assessed the impact on clinic attendance for eating disorder referrals.

Note that the examples in this table are from patient-facing DHAITs without examples of HCP-facing technologies.

As for economic evidence, the countries require cost-effectiveness analysis, based on quality-adjusted life years (QALYs). However, when looking at healthcare professional-facing DHAITs, using QALY is inappropriate, as the HCPs-facing DHAITs indirectly affect the QALY. Alternative to this analysis is cost-minimisation analysis, which is allowed by NICE and HAS for cases of clinical equivalence, as cost-minimisation analysis builds its evaluation purely on cost. Another analysis is cost-consequence analysis, which is conditionally accepted in the UK and France for tools delivering indirect benefits like workflow efficiency. Cost-consequence analysis presents costs and outcomes separately rather than aggregating them into a single measure, allowing for a more disaggregated evaluation of economic and clinical impacts. This method is particularly relevant for DHAITs that provide workflow efficiencies, administrative benefits, or systemic healthcare improvements that may not be fully captured through traditional cost-utility analyses [55].

It is worth noting that the HTA frameworks that informed these insights have historically prioritized clinical effectiveness, patient-reported outcomes, and real-world impact as core components of evidence for healthcare interventions [65]. While this focal point has allowed for patient-facing innovations to flourish in the health sector, it equally presents challenges when evaluating DHAITs primarily built to support the healthcare professional and not directly affecting the patient. Herein, it is worth again reflecting on the DiGA provisions as they recognize indirect patient outcomes (e.g., patient-relevant procedural improvements) as valid outcomes of DHAITs.

Evidence thresholds and uncertainty

Substantial equivalence is a regulatory evaluation method used to determine whether a new technology is similar to an already-approved device in terms of function, safety, and effectiveness. This approach is used primarily in the United States and South Korea, where AI-based medical devices can be approved without requiring extensive new clinical trials if they demonstrate equivalence to an existing cleared technology. The FDA's 510(k) pathway in the United States applies substantial equivalence assessments to AI-based medical devices, particularly radiology and diagnostics [66]. Under this framework, manufacturers must demonstrate that their AI-driven tool has similar technological characteristics and intended use as a previously approved comparator device. Substantial equivalence is not formally used in the policy landscapes of Germany and the UK. France uses the concept of substantial equivalence (in the form of technical and functional equivalence) to reimburse medical devices, though not for regulatory approval. In Germany's DiGA framework, digital health applications can be evaluated based on prior evidence of effectiveness, reducing the burden of additional clinical trials when modifications do not substantially alter functionality [67].

Furthermore, all six studied countries require sensitivity analyses as part of their evidence requirements to further limit the uncertainty surrounding the evidence provided. Currently, two countries explicitly mandate probabilistic approaches to sensitivity analyses:

- In the UK, NICE explains that all inputs used in economic evaluations are estimated with a degree of imprecision. The NICE health technology evaluations manual specifies that probabilistic sensitivity analysis is preferred for translating the imprecision in all input variables into a measure of decision uncertainty of the options being compared. In non-linear decision models, probabilistic methods provide the best estimates of mean costs and outcomes. NICE requires the mean value, distribution around the mean, and the source and rationale for the supporting evidence to be clearly described for each parameter included in the model [68].

- In France, HAS explains that a systematic exploration of the sources of uncertainty associated with the evaluation's structural choices, the modelling choices and the model parameter estimations should be presented according to an appropriate methodology [53]:
  - Sensitivity analyses should quantify the impact of a different structural choice in the reference-case analysis (e.g. perspective, time horizon, population analysed, comparators, discount rate).
  - Sensitivity analyses should quantify the impact of methodological choices and modelling assumptions (e.g., model structure, data sources, calculation methods, or assumptions to estimate the value of parameters not directly observed). The impact of the assumptions used for extrapolating treatment effects should be systematically explored.
  - The uncertainty associated with the parameters of the model should be systematically explored using two complementary approaches: a probabilistic sensitivity analysis, based on a second-order Monte Carlo simulation, and deterministic sensitivity analyses identifying the parameters (or combinations of parameters) that have the greatest influence on the results of the evaluation.

In France, HAS also allows using Bayesian models in their HTA pathway [52], particularly for technologies where real-world performance may differ from initial validation studies. Integrating Bayesian methods into HTA processes can offer a more robust framework for managing the inherent uncertainties associated with AI-based health interventions [69]. By combining prior knowledge with observed data, Bayesian approaches provide a coherent mechanism for updating beliefs about a technology's effectiveness as new evidence emerges. This dynamic updating process could be particularly valuable in the context of AI, where continuous learning and adaptation are fundamental characteristics. The flexibility of Bayesian methods allows for incorporating diverse data sources, facilitating more comprehensive evaluations of AI technologies. Moreover, Bayesian frameworks enable probabilistic interpretations of outcomes, offering decision-makers explicit quantifications of uncertainty.

Finally, in the context of AI technologies, it is vital to consider that they can evolve and adapt post-deployment. To accommodate this necessity, initial initiatives have been introduced outside of Europe and within the EU:

- Canada and South Korea implemented post-market surveillance and change approval requirements for modifications in AI systems that impact safety and effectiveness. It is worth noting that in South Korea, change approval and certification are exempt if additional training data is added to improve accuracy, as long as manufacturers manage both the training data and device performance through a quality management system [70].

- The United States has implemented Pre-Approved Controlled Change Plans (PCCPs), which allow pre-specified modifications to an AI model without requiring a new regulatory submission for each update. PCCPs establish performance thresholds and validation requirements before deployment, ensuring continued compliance with safety and efficacy standards [71].
- The newly implemented EU AI Act has a similar provision. It states that if a high-risk AI system is being modified substantially, meaning a change to the operating system and software which modifies the intended purpose, the system requires a new conformity assessment procedure. However, if the change to its system and performance have been pre-determined by the provider during the initial conformity assessment, this does not constitute a substantial modification. Furthermore, it requires providers to establish a post-market monitoring system for high-risk AI systems [72].

Policy Recommendations

Ultimately, the classification frameworks and evidence requirements for professional-facing DHAITs are in an embryonic state. Paired with the tremendous pace at which DHAITs are developed and can change over time, the policy landscape is under considerable pressure to identify quick and reliable solutions to profoundly complex problems. Based on the findings of this white paper, eight policy recommendations are established. These policy recommendations were drafted under the general philosophy that the rules and guardrails imposed by the policy landscape should foster purposeful and value-based innovation that benefits citizens, patients, health professionals, and health systems.

1) Develop functional classification frameworks aligned with risk profiles.

Issue: Current classification schemes conflate patient- and professional-facing tools and vary across jurisdictions, leading to mismatches between function and evidence requirements.

Recommendation: Establish clear functional classification systems that differentiate technologies based on intended use, user, and clinical risk, building on the taxonomy proposed in this report.

Rationale: A shared classification framework enables proportionate oversight and appropriate evaluation, serving as the backbone for consistent HTA and regulatory processes. It also allows clear mapping of technologies to the right evidence pathways.

Responsible actors:

- National regulators (e.g. FDA, European Medicines Agency, Medicines and Healthcare products Regulatory Agency [MHRA])
- HTA agencies (e.g. NICE, HAS, Canadian Agency for Drugs and Technologies in Health)
- International bodies (e.g. World Health Organization, World Economic Forum, IMDRF; specifically for harmonization efforts)
- Research teams and coalitions developing taxonomies (e.g. academic HTA centres, EUnetHTA, AssessDHT, EDiHTA)

2) Systematically link classification categories to tailored evidence standards and fit-for-purpose metrics.

Issue: Without a clear linkage between classification and evidence standards, assessment bodies risk applying inappropriate methods, such as requiring clinical trial data or health utility scores where system-level or workflow-focused metrics are more relevant. This issue extends beyond digital health and reflects broader challenges in applying utility measures in HTA.

Recommendation: Connect a purpose-driven classification framework with evidence requirements and evaluation metrics so that evidence expectations are proportionate to the intended use case.

Rationale: By mapping classification categories to suitable evidence standards and outcome metrics, HTA and regulatory bodies can ensure that evaluation methods are both proportionate and meaningful. This avoids underestimating the value of professional-facing tools, especially AI-driven diagnostics and CDS systems that do not directly affect QALYs but substantially improve system performance.

Responsible actors:

- HTA agencies (e.g. NICE, HAS, Canadian Agency for Drugs and Technologies in Health)
- Developers and health economists
- Regulators and classification standard-setters (e.g. IMDRF, national digital health bodies)

3) Prioritise Bayesian statistical methods for evaluating digital health and AI.

Issue: Traditional statistical methods cannot adequately model evolving AI systems or capture uncertainty in algorithm performance over time.

Recommendation: Integrate Bayesian methods into HTA frameworks to support continuous learning, real-world adaptation, and probabilistic reasoning.

Rationale: Bayesian approaches accommodate dynamic updates and enable more realistic assessments of benefit, supporting "living" HTAs that evolve alongside technologies.

Responsible actors:

- HTA agencies
- Academic methodologists and biostatistics groups
- Regulatory agencies to include Bayesian-compatible submission pathways

4) Implement Pre-Approved Controlled Change Plans for AI-based technologies.

Issue: Current regulatory frameworks are not well equipped to accommodate the iterative nature of AI systems, which require frequent updates to remain effective and safe. These updates often trigger time-consuming re-evaluation processes, creating regulatory bottlenecks and slowing innovation.

Recommendation: Adopt PCCPs to enable predefined, regulator-approved algorithmic modifications within clearly defined boundaries, ensuring updates remain within a validated performance envelope.

Rationale: Following the FDA model, PCCPs allow safe and efficient adaptation of AI systems by enabling specific, controlled modifications within a predetermined scope. A PCCP outlines which changes are permitted, how they should be implemented and monitored, and how their impacts will be assessed — ensuring that updates remain within a validated performance envelope and do not compromise safety or efficacy.

Responsible actors:

- National regulators (FDA, EMA, *Bundesinstitut für Arzneimittel und Medizinprodukte*, Health Canada)
- Standards bodies (International Organization for Standardization, Institute of Electrical and Electronics Engineers) to codify PCCP best practices
- Developers who must submit well-specified update plans

5) Align HTA, regulatory, and resource allocation processes through early stakeholder engagement and shared data standards.

Issue: Disjointed approval and reimbursement pathways lead to inefficiencies and increase time to access. The divide is especially stark for medical devices, where HTA often requires data types not generated for regulatory approval.

Recommendation: Facilitate early joint advice and develop mechanisms to align regulatory and HTA evidence requirements, drawing on practical examples such as the MHRA-NICE Early Value Assessment programme or DiGA's linkage of *Conformité Européenne* (CE)-marking to HTA eligibility.

Rationale: While complete convergence may not be realistic, coordinated scientific advice and alignment on data expectations can reduce duplication and make pathways more predictable. In some cases, harmonisation may require enhancing regulatory evidence expectations; in others, HTA frameworks may need to adapt to device-specific development models.

Responsible actors:

- Health ministries to mandate coordination structures
- HTA bodies and regulators to develop joint review mechanisms
- Developers to engage early through scientific advice

6) Broaden relevance HTA processes to non-patient-facing DHAITs by incorporating outcome metrics capturing system-level and professional-facing impacts, such as efficiency gains and workflow optimization.

Issue: Current HTA metrics such as QALYs overlook value generated by professional-facing technologies that impact workflow or system performance, bottlenecking innovation in these areas.

Recommendation: Expand the scope of HTA evaluations to include provider outcomes, operational efficiency, and decision-making quality, as outlined in the report's taxonomy.

Rationale: A broader evaluative lens ensures a full accounting of benefits, particularly for CDS tools, diagnostics, and administrative AI clinicians use.

Responsible actors:

- HTA agencies to revise evaluation guidance
- Payers to support broader value definitions
- Developers to generate relevant metrics beyond patient outcomes

7) Pilot and refine frameworks through multistakeholder engagement.

Issue: Frameworks developed without stakeholder input may be unworkable in practice or misaligned with clinical realities.

Recommendation: Co-design, pilot and regularly review regulatory, HTA, and resource allocation frameworks with clinicians, developers, payers, and other relevant stakeholders to ensure these frameworks continue to be aligned with the evolution of DHAITs.

Rationale: Co-design and piloting of frameworks promote feasibility, relevance, and trust. They also ensure frameworks remain adaptable to different contexts and stages of digital maturity. Furthermore, regular reviews are essential to ensure that frameworks remain in alignment with the development of the technology.

Responsible actors:

- Innovation units within ministries of health
- HTA and regulatory agencies
- Healthcare providers and specialty societies
- Industry consortia and patient organisations

8) Leverage international regulatory networks for regulating and evaluating healthcare professional facing DHAITs.

Issue: The fast-evolving field of DHAITs is outpacing the capacity of national regulatory and HTA bodies to develop suitable regulatory and evaluation frameworks. Fragmented regulation and evaluation of DHAITs leads to incoherent safety and quality standards, limiting trust, access, and innovation.

Recommendation: Leverage international regulatory networks, such as the IMDRF, for the regulation and evaluation of healthcare professional facing DHAITs.

Rationale: A regulatory network would support regulatory convergence, knowledge exchange, and consistent oversight across jurisdictions while maintaining local decision-making autonomy.

Responsible actors:

- HTA and regulatory agencies
- Standard bodies
- Industry consortia, patient organisations, and academia and clinical experts

References

1. World Health Organization. Nursing workforce grows, but inequities threaten global health goals [Internet]. Nursing workforce grows, but inequities threaten global health goals. 2025 [cited 2025 May 27]. Available from: https://www.who.int/news/item/12-05-2025-nursing-workforce-grows--but-inequities-threaten-global-health-goals

2. OECD. Health at a Glance 2023: OECD Indicators [Internet]. OECD; 2023 [cited 2025 May 1]. (Health at a Glance). Available from: https://www.oecd.org/en/publications/health-at-a-glance-2023_7a7afb35-en.html

3. National Centre for Rural Health and Care. Rural Workforce Issues in Health and Care - Executive Summary. 2018.

4. World Health Organization. A Global Health Strategy for 2025-2028: advancing equity and resilience in a turbulent world [Internet]. 2025. Available from: https://cdn.who.int/media/docs/default-source/documents/about-us/general-programme-of-work/global-health-strategy-2025-2028.pdf?sfvrsn=237faeeb_3

5. van Kessel R, Seghers LE, Anderson M, Monti G, Haig M, Schmidt J, et al. A scoping review and expert consensus on digital determinants of health. Bulletin of the World Health Organization. 2024;

6. Foreign, Commonwealt & Development Office (UK). Digital development strategy 2024 to 2030 [Internet]. 2024 [cited 2025 May 1]. Available from: https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030

7. World Health Organization. Going digital for noncommunicable disease: the case for action [Internet]. Geneva; 2024. Available from: https://iris.who.int/bitstream/handle/10665/378478/9789240089921-eng.pdf

8. World Health Organization. Classification of digital interventions, services and applications in health: a shared language to describe the uses of digital technology for health, 2nd ed [Internet]. [cited 2024 Jun 29] p. 66. Available from: https://www.who.int/publications/i/item/9789240081949

9. Lantzsch H, Panteli D, Martino F, Stephani V, Seißler D, Püschel C, et al. Benefit Assessment and Reimbursement of Digital Health Applications: Concepts for Setting Up a New System for Public Coverage. Front Public Health [Internet]. 2022 Apr 21 [cited 2025 Mar 15];10. Available from: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2022.832870/full

10. van Kessel R, Ranganathan S, Anderson M, McMillan B, Mossialos E. Exploring potential drivers of patient engagement with their health data through digital platforms: A scoping review. Int J Med Inform. 2024 Sep;189:105513.

11. Paccoud I, Leist AK, Schwaninger I, van Kessel R, Klucken J. Socio-ethical challenges and opportunities for advancing diversity, equity, and inclusion in digital medicine. Digit Health. 2024;10:20552076241277705.

12. European Commission. Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847 (Text with EEA relevance) [Internet]. Brussels: European Commission; 2025 Mar [cited 2024 Mar 13]. Available from: https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng

13. Segur-Ferrer J, Moltó-Puigmartí C, Pastells-Peiró R, Vivanco-Hidalgo RM. Methodological Frameworks and Dimensions to Be Considered in Digital Health Technology Assessment: Scoping Review and Thematic Analysis. Journal of Medical Internet Research. 2024 Apr 10;26(1):e48694.

14. World Health Organization. WHO guideline: recommendations on digital interventions for health system strengthening [Internet]. Geneva: World Health Organization; 2019 [cited 2025 Jan 9]. Available from: https://iris.who.int/handle/10665/311941

15. Schmidt J, Schutte NM, Buttigieg S, Novillo-Ortiz D, Sutherland E, Anderson M, et al. Mapping the regulatory landscape for artificial intelligence in health within the European Union. npj Digit Med. 2024 Aug 27;7(1):1–9.

16. Brönneke JB, Herr A, Reif S, Stern AD. Dynamic HTA for digital health solutions: opportunities and challenges for patient-centered evaluation. International Journal of Technology Assessment in Health Care. 2023 Jan;39(1):e72.

17. Haig M, Main C, Chávez D, Kanavos P. A Value Framework to Assess Patient-Facing Digital Health Technologies That Aim to Improve Chronic Disease Management: A Delphi Approach. Value in Health. 2023 Oct 1;26(10):1474–84.

18. Gomes M, Murray E, Raftery J. Economic Evaluation of Digital Health Interventions: Methodological Issues and Recommendations for Practice. Pharmacoeconomics. 2022;40(4):367–78.

19. Project – EDiHTA [Internet]. [cited 2025 May 28]. Available from: https://edihta-project.eu/project/

20. ASSESS DHT [Internet]. 2024 [cited 2025 May 28]. Available from: https://assess-dht.eu/

21. Digital Therapeutics Alliance, Wade B, Abraham J, Coder M. Guidance to the Industry: Classification of Digital Health Technologies.

22. Boers M, Rochereau A, Stuwe L, Miguel LS, Klucken J, Mezei F, et al. Classification grid and evidence matrix for evaluating digital medical devices under the European union landscape. npj Digit Med. 2025 May 24;8(1):1–10.

23. OECD. Health in the 21st Century: Putting Data to Work for Stronger Health Systems [Internet]. Paris: OECD Publishing; 2019 Nov [cited 2025 Mar 15]. Available from: https://doi.org/10.1787/e3b23f8e-en.

24. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. The Lancet Digital Health. 2021 Mar 1;3(3):e195–203.

25. Muehlematter UJ, Bluethgen C, Vokinger KN. FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks. The Lancet Digital Health. 2023 Sep 1;5(9):e618–26.

26. Aboy M, Minssen T, Vayena E. Navigating the EU AI Act: implications for regulated digital medical products. npj Digit Med. 2024 Sep 6;7(1):1–6.

27. Palmieri S, Goffin T. A Blanket That Leaves the Feet Cold: Exploring the AI Act Safety Framework for Medical AI. European Journal of Health Law. 2023 Feb 7;30(4):406–27.

28. European Commission. Overview Medical Device Sector [Internet]. 2025 [cited 2025 Mar 15]. Available from: https://health.ec.europa.eu/medical-devices-sector/overview_en

29. Jain A, Way D, Gupta V, Gao Y, de Oliveira Marinho G, Hartford J, et al. Development and Assessment of an Artificial Intelligence-Based Tool for Skin Condition Diagnosis by Primary Care Physicians and Nurse Practitioners in Teledermatology Practices. JAMA Netw Open. 2021 Apr 1;4(4):e217249.

30. Gagliardi KS, Coleman S, Intinarelli G, Karliner L, Appelle N, Taylor B, et al. An Automated Telephone Call System Improves the Reach and Cost-effectiveness of Panel Management Outreach for Cancer Screening. The Journal of Ambulatory Care Management. 2020 Jun;43(2):148.

31. Niewada M, Tabor B, Piotrowicz E, Piotrowicz R, Opolski G, Banach M, et al. Cost-effectiveness of telerehabilitation in patients with heart failure in Poland: an analysis based on the results of Telerehabilitation in the Heart Failure Patients (TELEREH-HF) randomized clinical trial. Kardiol Pol. 2021;79(5):510–6.

32. Lima HDJ, Souza RDSP, Santos ASMSE, Borges DL, Guimaraes ARF, Ferreira GVDBA, et al. Virtual reality on pulmonary function and functional independence after coronary artery bypass grafting: clinical trial. Am J Cardiovasc Dis. 2020;10(4):499–505.

33. Scodari BT, Chacko S, Matsumura R, Jacobson NC. Using machine learning to forecast symptom changes among subclinical depression patients receiving stepped care or usual care. J Affect Disord. 2023 Nov 1;340:213–20.

34. Ericson O, Hjelmgren J, Sjövall F, Söderberg J, Persson I. The Potential Cost and Cost-Effectiveness Impact of Using a Machine Learning Algorithm for Early Detection of Sepsis in Intensive Care Units in Sweden. J Health Econ Outcomes Res. 2022;9(1):101–10.

35. Patibandla S, Kumar K, Adepu R, Bandaru RK, Maduri B. The Validation of a Mobile Based Ambulatory Heart Rhythm Monitoring Solution - Vigo Heart. European Journal of Cardiovascular Medicine. 2024 Mar 20;14:374–85.

36. Chang J, Bliss L, Angelov N, Glick A. Artificial intelligence-assisted full-mouth radiograph mounting in dental education. Journal of Dental Education. 2024;88(7):933–9.

37. Malcolm R, Shore J, Stainthorpe A, Ndebele F, Wright K. Economic evaluation of a vision-based patient monitoring and management system in addition to standard care for adults admitted to psychiatric intensive care units in England. J Med Econ. 2022;25(1):1101–9.

38. Jain S, Gupta M, Agarwal M, Jain P. Advances in Diabetic Retinopathy: Pioneering Technologies Transforming Patient Care. 2024;

39. Shaukat A, Lichtenstein DR, Somers SC, Chung DC, Perdue DG, Gopal M, et al. Computer-Aided Detection Improves Adenomas per Colonoscopy for Screening and Surveillance Colonoscopy: A Randomized Trial. Gastroenterology. 2022 Sep;163(3):732–41.

40. Xie Y, Nguyen QD, Hamzah H, Lim G, Bellemo V, Gunasekeran DV, et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. Lancet Digit Health. 2020 May;2(5):e240–9.

41. Wilson PM, Ramar P, Philpot LM, Soleimani J, Ebbert JO, Storlie CB, et al. Effect of an Artificial Intelligence Decision Support Tool on Palliative Care Referral in Hospitalized Patients: A Randomized Clinical Trial. J Pain Symptom Manage. 2023 Jul;66(1):24–32.

42. Blomberg SN, Christensen HC, Lippert F, Ersbøll AK, Torp-Petersen C, Sayre MR, et al. Effect of Machine Learning on Dispatcher Recognition of Out-of-Hospital Cardiac Arrest During Calls to Emergency Medical Services: A Randomized Clinical Trial. JAMA Netw Open. 2021 Jan 4;4(1):e2032320.

43. Mertens S, Krois J, Cantu AG, Arsiwala LT, Schwendicke F. Artificial intelligence for caries detection: Randomized trial. Journal of Dentistry. 2021 Dec 1;115:103849.

44. Bhattacharyay S, Caruso PF, Åkerlund C, Wilson L, Stevens RD, Menon DK, et al. Mining the contribution of intensive care clinical course to outcome after traumatic brain injury. npj Digit Med. 2023 Aug 21;6(1):1–11.

45. Balloch J, Sridharan S, Oldham G, Wray J, Gough P, Robinson R, et al. Use of an ambient artificial intelligence tool to improve quality of clinical documentation. Future Healthcare Journal. 2024 Sep 1;11(3):100157.

46. Shah D, Dewan L, Singh A, Jain D, Damani T, Pandit R, et al. Utility of a smartphone assisted direct ophthalmoscope camera for a general practitioner in screening of diabetic retinopathy at a primary health care center. Indian J Ophthalmol. 2021 Nov;69(11):3144–8.

47. Reynolds MR, Stein AB, Sun X, Hytopoulos E, Steinhubl SR, Cohen DJ. Cost-Effectiveness of AF Screening With 2-Week Patch Monitors: The mSToPS Study. Circ Cardiovasc Qual Outcomes. 2023 Nov;16(11):e009751.

48. Yau YW, Li Z, Chua MT, Kuan WS, Chan GWH. Virtual reality mobile application to improve videoscopic airway training: A randomised trial. Ann Acad Med Singap. 2021 Feb;50(2):141–8.

49. Xiang H, Shen J, Wheeler KK, Patterson J, Lever K, Armstrong M, et al. Efficacy of Smartphone Active and Passive Virtual Reality Distraction vs Standard Care on Burn Pain Among Pediatric Patients: A Randomized Clinical Trial. JAMA Network Open. 2021 Jun 21;4(6):e2112082.

50. Park A, Chute C, Rajpurkar P, Lou J, Ball RL, Shpanskaya K, et al. Deep Learning–Assisted Diagnosis of Cerebral Aneurysms Using the HeadXNet Model. JAMA Network Open. 2019 Jun 7;2(6):e195600.

51. National Institute for Health and Care Excellence. Evidence standards framework for digital health technologies [Internet]. London: National Institute for Health and Care Excellence; 2022 Aug [cited 2025 Mar 17]. Available from: https://www.nice.org.uk/corporate/ecd7

52. Haute Autorité de Santé. Methodological Choices for the Clinical Development of Medical Devices [Internet]. Saint-Denis: Haute Autorité de Santé; 2021 Jun [cited 2025 Feb 10]. Available from: https://www.has-sante.fr/jcms/c_1696842/en/methodological-choices-for-the-clinical-development-of-medical-devices

53. Haute Autorité de Santé. Choices in Methods for Economic Evaluation [Internet]. Saint-Denis: Haute Autorité de Santé; 2020 Jul [cited 2025 Mar 17]. Available from: https://www.has-sante.fr/jcms/r_1499251/en/choices-in-methods-for-economic-evaluation

54. Lau K, Halligan J, Fontana G, Guo C, O'Driscoll FK, Prime M, et al. Evolution of the clinical simulation approach to assess digital health technologies. Future Healthc J. 2023 Jul;10(2):173–5.

55. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions—A call for innovative evidence generation approaches. npj Digit Med. 2020 Aug 27;3(1):1–14.

56. Tan R, Kavishe G, Kulinkina AV, Renggli S, Luwanda LB, Mangu C, et al. A cluster randomized trial assessing the effect of a digital health algorithm on quality of care in Tanzania (DYNAMIC study). PLOS Digit Health. 2024 Dec 23;3(12):e0000694.

57. Clermont G, Bartels J, Kumar R, Constantine G, Vodovotz Y, Chow C. In silico design of clinical trials: a method coming of age. Crit Care Med. 2004 Oct;32(10):2061–70.

58. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps. Annals of Behavioral Medicine. 2019 May 3;53(6):573–82.

59. Blonde L, Khunti K, Harris SB, Meizinger C, Skolnik NS. Interpretation and Impact of Real-World Clinical Data for the Practicing Clinician. Adv Ther. 2018;35(11):1763–74.

60. Aziz MT, Ur-Rehman T, Qureshi S, Bukhari NI. Reduction in Chemotherapy Order Errors with Computerised Physician Order Entry and Clinical Decision Support Systems. HIM J. 2015 Oct;44(3):13–22.

61. Aziz AFA, Ong T. Real-World Outcomes of a Digital Behavioral Coaching Intervention to Improve Employee Health Status: Retrospective Observational Study. JMIR mHealth and uHealth. 2024 Sep 10;12(1):e50356.

62. Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. JAMA. 2022 Dec 27;328(24):2446.

63. Simon GE, Shortreed SM, DeBar LL. Zelen design clinical trials: why, when, and how. Trials. 2021 Aug 17;22(1):541.

64. Denison-Day J, Muir S, Newell C, Appleton KM. A Web-Based Intervention (MotivATE) to Increase Attendance at an Eating Disorder Service Assessment Appointment: Zelen Randomized Controlled Trial. Journal of Medical Internet Research. 2019 Feb 27;21(2):e11874.

65. Kristensen FB, Husereau D, Huić M, Drummond M, Berger ML, Bond K, et al. Identifying the Need for Good Practices in Health Technology Assessment: Summary of the ISPOR HTA Council Working Group Report on Good Practices in HTA. Value Health. 2019 Jan;22(1):13–20.

66. US Food and Drugs Administration. The 510(k) Program: Evaluating Substantial Equivalence in Premarket Notifications [Internet]. Rockville, MD: US Food and Drugs Administration; 2019 Mar [cited 2025 Mar 17]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/510k-program-evaluating-substantial-equivalence-premarket-notifications-510k

67. Bundesministerium für Gesundheit. Verordnung über das Verfahren und die Anforderungen zur Prüfung der Erstattungsfähigkeit digitaler Gesundheitsanwendungen in der gesetzlichen Krankenversicherung [Internet]. 2020 Apr. Available from: https://www.gesetze-im-internet.de/digav/BJNR076800020.html

68. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual [Internet]. London, UK: National Institute for Health and Care Excellence; 2023 Oct [cited 2024 Dec 8]. Available from: https://www.nice.org.uk/process/pmg36/chapter/economic-evaluation-2

69. Voets MM, Veltman J, Slump CH, Siesling S, Koffijberg H. Systematic Review of Health Economic Evaluations Focused on Artificial Intelligence in Healthcare: The Tortoise and the Cheetah. Value in Health. 2022 Mar 1;25(3):340–9.

70. Ministry of Food and Drug Safety. Guidance on the Review and Approval of Artificial Intelligence (AI)-based Medical Devices [Internet]. 2023 Jul. Available from: https://www.mfds.go.kr/eng/brd/m_40/view.do?seq=72627

71. US Food and Drugs Administration. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions [Internet]. Rockville, MD: US Food and Drugs Administration; 2024 Dec [cited 2025 Feb 15]. Available from: https://www.fda.gov/media/166704/download

72. European Commission. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence [Internet]. PE/24/2024/REV/1 Jul 12, 2024. Available from: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

LSE Consulting

+44 (0)20 7107 5215
consulting@lse.ac.uk
lse.ac.uk/consultancy

LSE Consulting
Houghton Street
London WC2A 2AE