ICER-PHTI Assessment Framework for Digital Health Technologies

September 2023

Table of Contents

List of Acronyms and Abbreviations Used in this Framework

AI	Artificial Intelligence
BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte
CBT	Cognitive behavioral therapy
DHTs	Digital Health Technologies
3D	Three-dimensional
EMA	European Medical Agency
EU	European Union
FDA	Food and Drug Administration
HTA	Health Technology Assessment
IMDRF	International Medical Device Regulators Forum
IT	Information Technology
mHealth	Mobile health
MHRA	Medicines and Healthcare products Regulatory Agency
NICE	National Institute for Health and Care Excellence
RCT	Randomized Controlled Trial
SaMD	Software as a Medical Device
SiMD	Software in a Medical Device
TGA	Therapeutic Goods Administration

Overview

This value assessment framework describes the conceptual model and associated methods that will guide assessments of digital health technologies (DHTs) within broader evaluative reports produced by the Peterson Health Technology Institute (PHTI). PHTI aims to produce publicly available, objective, rigorous evaluations of DHTs to accomplish two goals: to set evidence standards that guide technology developers to generate robust evidence on their products; and to provide reviews that help organizations adopt high-impact DHTs with the strongest evidence for delivering improved clinical outcomes and cost savings.

Given the unique nature of DHTs, existing value assessment frameworks and evidence standards for health technologies such as drugs and devices are not directly applicable. All stakeholders will benefit from comprehensive and explicit standards of evidence on the different dimensions necessary to understand the value of DHTs. This framework therefore is intended to describe the standards of evidence that DHT developers should meet when designing the early studies of their products, and to delineate how that evidence will be assessed in PHTI DHT evaluations.

Background

DHTs are a broad and rapidly innovating class of health technology with distinctive pathways for development, regulatory approval, uptake, and reimbursement. There is no single uniform definition or taxonomy for DHTs. The Food and Drug Administration (FDA) presents their definition of DHTs in the following terms:

Digital health technologies use computing platforms, connectivity, software, and sensors for health care and related uses. These technologies span a wide range of uses, from applications in general wellness to applications as a medical device. They include technologies intended for use as a medical product, in a medical product, as companion diagnostics, or as an adjunct to other medical products (devices, drugs, and biologics).

According to the FDA, this definition encompasses categories as disparate as mobile health (mHealth), health information technology (IT), wearable devices, and telehealth. When software and apps are integrated with traditional medical devices, there is existing regulatory guidance to determine whether a product should be considered a standalone DHT or if it should be evaluated as a traditional medical device. To help make this distinction, the International Medical Device Regulators Forum (IMDRF) and the FDA define Software *as* a Medical Device (SaMD) as "software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device." In other words, to be considered as SaMD, the software needs to function independently of the existing medical device. In contrast, if the device cannot be used without the software, whether it is running the mechanics or processing the information that is produced, it is considered Software *in* a Medical Device (SiMD). Only SaMD will be considered within the scope of this value framework since we believe that SiMD can be evaluated using traditional approaches to the assessment of medical devices.

SaMD and other DHTs span multiple broad functional categories such as communication, monitoring, diagnosis, prevention, disease management, and treatment. Within these broad functional categories, DHTs can still vary significantly with different goals, end-users (e.g., patient vs. clinician), health risk to users, and financial risk to

payers and other purchasers. The primary goal of some DHTs is to reduce inefficiencies in care pathways or improve access to effective care, whereas others are designed to make a clinical diagnosis, give prognostic information that guides selection of therapy, or directly deliver therapeutic interventions. Clarity on goals and end-users is critical to understanding the level of risk inherent in DHTs and corresponding evidentiary standards for effectiveness and value.

Regulators, health technology assessment (HTA) bodies, and payers around the globe have taken different approaches to adapting their assessments to meet the evolving field of DHTs (see Table 1). There is no uniformity across country-specific definitions of DHTs nor in the requirements that determine whether a DHT is subject to standard medical device regulatory pathways. Several regulatory bodies—including the FDA—are developing new regulatory pathways and/or specific guidance on DHTs but have not completed their efforts. The FDA has recently published draft guidance for DHT developers on the elements of information required for premarket submission, and has established a Digital Health Center of Excellence to provide regulatory advice and support for the review of DHTs. Despite these early efforts, the pathways for DHT development, potential regulatory approval, and evaluation for implementation and payment remain fluid and inconsistent, leading to the potential for misalignment between the evidence generated by developers to support DHT uptake and the evidence judged necessary to demonstrate value by health systems and payers. This misalignment can lead to the dismissal of DHTs that could improve outcomes and lower costs while also risking the adoption of some DHTs that do neither.

Given the diversity of DHT modalities and the lack of standard regulatory and payer pathways, it is not surprising that no clear set of evidentiary standards by which to evaluate DHTs has yet emerged. Across the sample of regulatory and HTA agencies examined, the overarching factor that determines regulatory and payer evidentiary requirements for DHTs is their function and, ultimately, the risk posed to patients. The FDA draft guidance separates DHTs into three categories depending upon the level of "concern" given the risks to patients. However, the FDA does not suggest standards for the type of studies or evidence that is required. In the UK, the National Institute for Health and Care Excellence (NICE) has established a set of risk-based tiers to guide their assessment of DHTs for the UK National Health Service. The ICER-Peterson assessment framework also proposes a risk-based model and puts forth a novel approach to evaluate cost savings. Notable, our initial framework positions evaluation of clinical effectiveness and economic impact as two primary domains whose outcomes are informed by other attributes of the technology, including user experience, impact on health equity, privacy, and data security.

The use of artificial intelligence and machine learning in digital health software presents a unique assessment challenge, noted by the FDA in preliminary guidance published in 2021. By their nature, AI algorithms are difficult to evaluate and the quality and accuracy of the algorithms' predictive power depends heavily on the details of the training dataset and model parameters, which may be difficult to assess by an outside entity. Future versions of this framework will more specifically address how to assess the unique aspects of AI within DHTs.

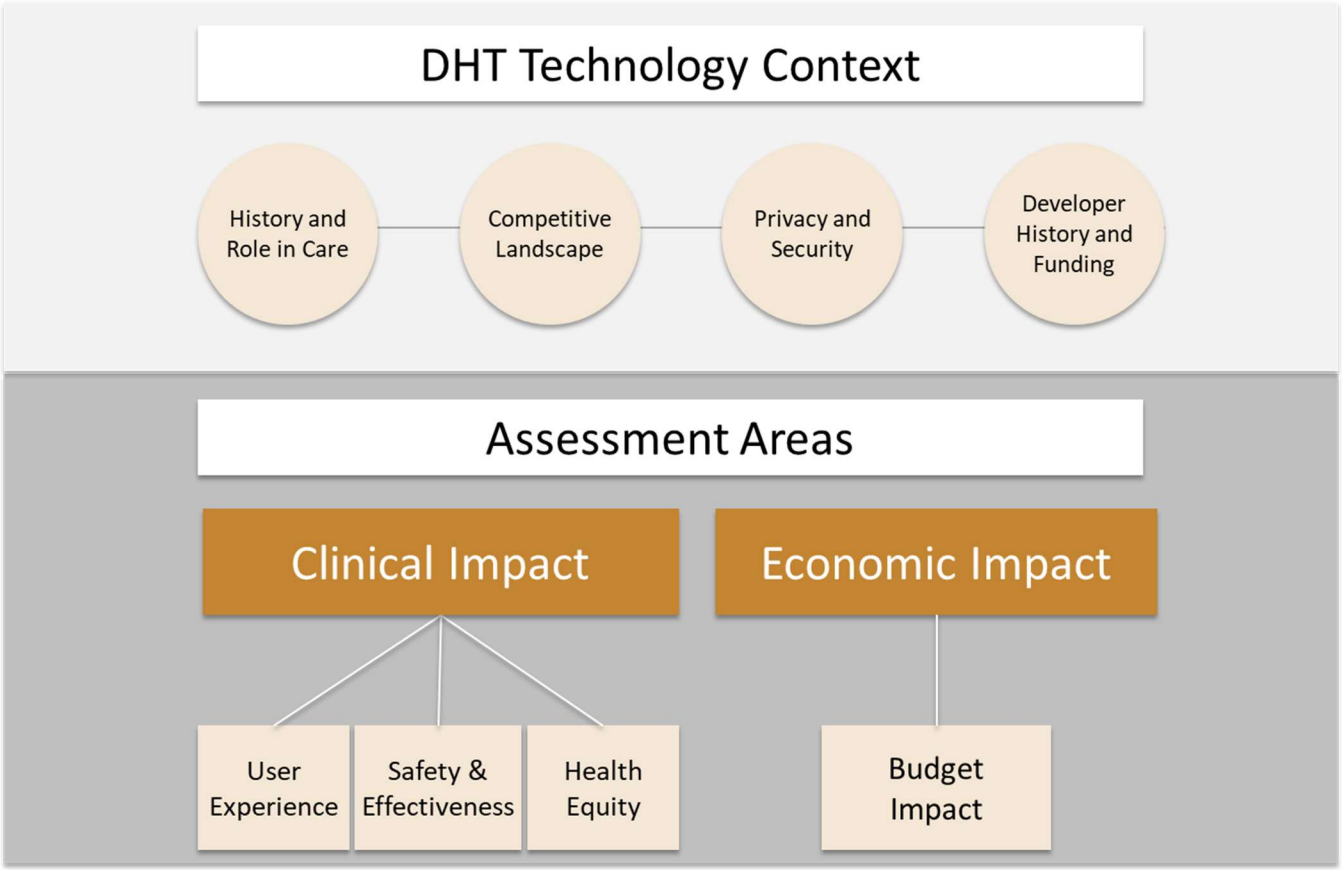
Table 1. Overview of various approaches to DHT assessment by a sample of regulatory bodies, HTA agencies, and other stakeholders				
Regulatory Body (Locale)	**Scope**	**Separate Regulatory Path?**	**Guidelines on assessment of DHTs?**	**Overview**
FDA (US)	"Mobile health (mHealth), health information technology (IT), wearable devices, telehealth and telemedicine, and personalized medicine"	In development	In development	The Digital Health Center of Excellence of the FDA has published draft guidelines and resources on the regulator aspects of DHTs. (Updated Feb 2022)
EMA (EU)	"Digital technology-based methodologies to support approval of medicinal products." Examples include, but not limited to: sensors, wearables, mHealth, digital analytics.	No	Yes	The EMA has published a document providing general guidance that is not intended to be comprehensive. However, it does provide high level information for evidence standards for some digital technologies (see Section 6 of the linked document).
Health Canada (Canada)	Wireless Medical Devices, Mobile Apps, Telehealth, SaMD, AI	In development	No	Health Canada is establishing a new division for premarket review of digital health technologies, but detailed information is not available at the time of the initial release of this report. Health Canada last released an updated statement in May 2019.
MHRA (UK)	Apps, stand-alone software, or diagnostic devices that gather data from a person, include: SaMD, SiMD, AI	In development	In development	In September 2021, the MHRA released a framework for the development of regulatory pathways and evidence standards for SaMD and SiMD.
TGA (Australia)	SaMD Others	No	In development	TGA has published information on regulation that applies to software and apps which meet the legislated definition of a medical device in Australia.
BfArM (Germany)	Digital health applications (DiGA) are DHTs used by a patient or provider with a medical purpose: recognition, monitoring, treatment or alleviation of disease. Does not cover prevention.	Yes, rapid review process for qualified DiGA	Yes	Scope and evidence standards are published and include general recommendations on types on study designs with the goal of demonstrating a positive effect on health.
HTA Agency				
NICE (UK)	Broad categorization of digital health technologies, but excludes those that rely on AI with adaptive algorithms	N/A	Yes	NICE has evidentiary standards for DHTs based on risk and function
Agencia de Calidad Sanitaria de Andalucía (Spain)	Mobile Health Apps	N/A	Yes, but only mobile health apps are addressed	The Agencia de Calidad Sanitaria de Andalucía Complete has released a list of recommendations on design, use and assessment of health apps.
Other				
IMDRF (International)	SaMD, SiMD	N/A	Yes	IMDRF has published guidance based on risk categorization – "Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations."
European Commission (EU)	SaMD, SiMD depending on the intended use of the product (e.g., to improve health)	N/A	Yes	The EU published guidance on mHealth applications only.

AI: Artificial Intelligence, BfArM:Bundesinstitut für Arzneimittel und Medizinprodukte, EMA: European Medical Agency; EU: European Union, FDA: Food and Drug Administration, IMDRF: International Medical Device Regulators Forum, IT: Information technology, mHealth: mobile health, MHRA: Medicines and Healthcare products Regulatory Agency, NICE: National Institute for Health and Care Excellence, SaMD: Software as a Medical Device, SiMD: Software in a Medical Device, TGA: Therapeutic Goods Administration

Overall Conceptual Framework

In line with ICER's existing Value Assessment Framework for drugs, assessment of the value of DHTs requires consideration of multiple domains, as shown in Figure 1, including a descriptive domain on "DHT Technology Context" and two assessment area domains: "Clinical Impact," and "Economic Impact." Description of these domains, evidence standards for different types of DHTs based on risk categorization, and core methods for assessment of each domain are provided in sections below.

Figure 1. ICER-PHTI Assessment Framework for DHTs



Technology Context

Evaluation of the value of a DHT requires a clear understanding of the context in which the technology has been developed, including 1) the history of the DHT's development and its envisioned role in care, including proposed mechanisms for coding, billing, or other payment mechanisms; 2) the competitive landscape; 3) the status of its evaluation for privacy and security features; and 4) the developer's history and status of funding. Each of these elements of Technology Context are important in the overall value assessment of a DHT and will be developed as qualitative sections of a final report without separate ratings or inclusion as part of the assessment of clinical and economic impact. PHTI and their assessment partners will need to assure a consistent, rigorous approach to presenting this information.

DHT History and Role in Care, Competitive Landscape, Privacy and Security, and Developer History and Funding

History and Role in Care. The history of the development of the DHT is an important first element in an assessment. The DHT may be the first iteration of its type, on the cusp of rapid evolution with feedback from end-users. Or it may be less likely to evolve, the product of multiple cycles of adaptation across diverse patients and practice settings. The adaptability of the platform delivering a DHT and/or its ability to integrate with other IT systems is also important to describe as part of the technology context. Some DHTs may address a single clinical issue at launch, e.g., diet, but there may already be plans to broaden the scope to include other goals, such as medication use, or even additional clinical areas. It is also important to determine whether the DHT is provided through an IT platform that has distinctive advantages or disadvantages with integration alongside existing certain electronic health record or other IT systems.

Along with the history of the DHT, an analysis of its proposed role in care is central to consideration of its value. This analysis will cover the clinical and/or system problem for which the DHT is a potential solution. The scale of unmet need and current evidence on inefficiencies or gaps in care will be explored from the perspective of the end-user and health system. The analysis of the role in care will also include consideration of options and formal proposals for billing codes and overall payment arrangements, including fee-for-service, outcomes-based agreements, and subscription models.

Competitive Landscape. Decision-makers also need to understand the competitive landscape in which a DHT is emerging. Knowledge of the relative positioning of other companies and products in the same space will help inform decisions regarding whether to consider waiting for adoption until the DHT can be fully compared to other products at comparable stages of evolution.

Privacy and Security. A description of the status of privacy and security status of the DHT is an essential part of the context regarding the technology. The term "privacy" here refers to the ability that DHT end-users should have to control how their personal health information is stored and shared.

Patients can benefit from accessing their own health data, as well as from the increased efficiency of providers who share data in a manner compliant with the Health Insurance Portability and Accountability Act (HIPAA) and other regulations. However, as most DHT developers do not fall within HIPAA's purview, the ways in which a developer ensures the protection of personal health information must be evaluated outside of the federal framework for traditional health care providers.

Other third-party organizations have emerged to provide guidelines, tools, and certifications specific to the digital health space. For example the Health Information Trust Alliance (HITRUST), a company that has developed a series of validated security assessments and corresponding certifications with degrees of rigor that vary based on the organization's risk profile. Another example is the Organization for the Review of Health and Care Apps (ORCHA) is a private consultancy that performs assessments of DHTs, with data security and privacy as one of the four main topic areas for assessment. The ORCHA data and privacy assessment incorporates standards set by other assessment entities and looks at data use and data storage and transit, as well as the privacy information that is publicly available for the end-user.

Developer History and Funding. Lastly, purchasers and payers need to understand the history and funding of developers of DHTs because of the heightened risk of company failure in the ecosystem of DHTs. Key elements of this assessment include the longevity of the developer and history of its senior leadership, their track record for bringing DHTs into wide use, prior adoption by health systems and payment by insurers, and history and scale of venture capital and other investor funding.

Clinical Impact

User Experience

Broadly, user experience, often also called usability, can be thought of as the ease and satisfaction with which end-users can use the DHT for its intended purpose. The user experience drives end-user engagement and consistent use of DHTs, and, consequently, user experience is an important component of the evidence that should be developed and assessed for emerging DHTs. The terms user experience and usability are the most widely used in a family of similar ideas, including acceptability, satisfaction, quality, feasibility, and engagement. Although the field is still developing consensus on the best ways to measure this domain, there are established and validated instruments to measure usability, such as the System Usability Scale. Whether DHT developers use this instrument or another validated tool, they should strive to use the same measure(s) that have been previously used to evaluate similar DHTs to allow for robust indirect comparisons. Engagement of end users in the design and development process should be clearly documented.

User experience can be measured among prospective end-users, whether in formative stages (e.g., early demonstration of a product) or as part of a later-stage trial or study. Two general requirements should be noted. First, user experience should be measured among a diverse set of prospective or active end-users. The goal should be to reflect the anticipated real-world population of end-users, whether clinicians or patients. The population evaluated should not be highly selective in terms of age, gender, levels of comorbidities, language, or other factors that may be influential in the user experience with the product. Second, for those DHTs that will be used within a health care setting, user experience should be measured under a diverse set of practice settings and conditions. For example, the user experience with a prognostic DHT implemented within an integrated delivery system may differ widely from that in a more diffuse practice network. Single site studies with carefully cultivated care providers or patients should not serve as the basis for demonstrating user experience.

The evidence on user experience on an individual DHT will be assessed for its strengths and limitations, with key uncertainties, including generalizability, highlighted for decision-makers. When the assessment compares multiple

DHTs, user experience will also be framed in a comparative manner, with user experience for a DHT versus its comparator rated as: 1) Inferior; 2) Comparable; 3) Small Superiority; 4) Moderate-Large Superiority.

Safety and Effectiveness: DHT Functional Categories and Evidence Tiers

Assessing the safety and effectiveness of DHTs presents distinctive challenges given the wide array of health technology that is considered a DHT. Chief among these concerns is that the function of DHTs can vary considerably and this plays a direct role in determining the level of evidence that is required to judge the balance of benefits and risks with an appropriate level of certainty. The ICER-Peterson framework outlines three broad functional categories of DHTs to guide the assessment of safety and effectiveness.

These functional categories are shown below in Table 2 on the following page: (1) Self-Directed Health Management; (2) Professionally Directed Diagnostic and Prognostic Health Management; and (3) Professionally Directed Therapeutic Health Management. The categories are not hierarchical; each reflects different functions with associated risk levels that will be linked to distinct evidence standards in the following section. DHTs that perform Administrative Health Functions, including electronic health records and electronic prescribing platforms, are outside the scope of the first version of this framework.

Each functional category is then stratified based on its potential risk to patients across three qualitative categories: low risk, moderate risk, and high risk. The risk rating is a combination of the probability and the magnitude of potential harms that might result from the DHT during routine use. Although there are no quantitative thresholds proposed to distinguish the different risk levels, general framing and calibration are possible. The lowest level of risk is associated with DHTs that do not affect clinical care and are either meant to inform individuals about their health in a way that is not intended to be immediately actionable by a clinician, or that help provide health behavior management for prevention or for a condition that is not urgent. At the upper bounds of risk are DHTs that deliver an intervention or directly impact clinical decision-making for serious or urgent conditions, situations in which there is the potential for serious morbidity or mortality with inappropriate, incorrect, or ineffective use.

Table 2. Overview of Evidence Tiers for Assessment of DHT Safety and Effectiveness

Broad Functional Category	Potential Risk to Patients	Evidence Tier	Functions	Examples
Self-Directed Health Management	Low Risk	Tier 1	Personalized health information for use by the end-user not intended for professional consideration	Mobile health applications available for individual use without clinician involvement. Personalized health information for home use, electronic diaries, or risk assessment tools that can be delivered via Cloud, Internet, or App. Wearables and other "smart" devices for personal use (e.g., fitness trackers, apps that utilize smart watches)
Professionally-Directed Diagnostic and Prognostic Health Management	Moderate Risk to High Risk	Tier 2	Diagnoses a specific clinical condition and/or guides diagnosis or management decisions through diagnosis or prognosis. DHT use is directed by a medical professional or provides information that would be utilized in consultation with a medical professional.	DHTs with active monitoring that automatically records data and transmits it directly to healthcare professional for clinical decision making (e.g., sensor worn on the body), or monitoring for potentially serious conditions. DHTs in this category may provide immediate feedback to end-user but have the potential to trigger consultation with clinicians (e.g., App for atrial fibrillation monitoring).
Professionally-Directed Preventative and Therapeutic Health Management	Low Risk to Moderate Risk	Tier 3a	Preventative health behavior management with professional involvement	Prescribed behavior change technologies (e.g., smoking cessation, weight loss, insomnia.)
	Moderate Risk to High Risk	Tier 3b	Therapeutic. Directly provides treatment or acts as an adjunct to other interventions for a diagnosed clinical condition	Mobile applications that deliver a therapeutic intervention (e.g., CBT for behavioral health); DHTs that guide treatment or medical interventions such as wearables that detect periods of apnea during sleep and alarms to rouse the person. Software that creates 3D reconstruction images and determines location

CBT: Cognitive behavioral therapy, DHT: digital health technology, 3D: three-dimensional

Safety and Effectiveness: Evidence Standards and Evaluation

As noted earlier, the evidentiary standards in each tier for demonstrating safety and effectiveness are based on the potential function of the DHT and its relative risk to end-users. Table 3 on the following page describes the factors that help evaluate whether the risk is relatively low or high within each tier and corresponding "minimum" and "best" evidence standards for the tier.

Ideally, developers should have extensive discussions with relevant health providers, plan sponsors, and payers to agree on the level of risk and the corresponding evidence requirements as early as possible in the process of developing an evidence dossier for an emerging DHT. Additionally, developers should seek guidance on the most relevant clinical and economic outcomes early in the process. Outcomes should be relevant to both the user (e.g., patient health outcomes) and the payer (e.g., treatment engagement, health care utilization, clinical endpoints).

There are several factors that suggest the relative risk within the general range for each DHT functional category. DHTs that represent the first attempt to apply that modality or that are the first to use the general medium carry greater uncertainty regarding both benefits and potential harms, and therefore are likely to be viewed as having greater risk for patients. The most important factor, however, is the clinical consequence of obtaining inaccurate information or of having the delivery of the intervention fail to achieve its purpose. The more serious the consequences of DHT "failure," the more likely that the highest evidence standard within the functional category should be the goal of evidence development to support assessment of the DHT.

Minimum evidence standards, which correspond with lower relative end-user risk within the tier, should be viewed as the absolute minimum evidence needed for the assessment of the clinical effectiveness of a DHT. These standards have been selected to be feasible for developers but will, in most cases, represent a higher level of evidence than has been the norm in the current DHT market in the US. There will always be a tension between the desire for better evidence and the costs and feasibility of generating that evidence. The proposed "best" evidence standards seek to strike an appropriate balance and to set a reasonable bar for DHTs which have the potential to harm end-users. Further description of each set of evidence standards is provided after Table 3 below.

Table 3. Minimum and Best Evidence Requirements of Safety and Effectiveness for Each DHT Evidence Tier				
Broad Functional Category	**Evidence Tier**	**Contextual Factors to Determine Risk and Minimum vs. Best Evidence Requirement**	**Minimum Evidence Requirement**	**Best Evidence Requirement**
Self-Directed Health Management	Tier 1	Not applicable	The DHT provides valid, accurate, up to date information.	Empirical data from users on *perceived* behavior change or health improvement, typically a pre- / post- study design within users.
Professionally Directed Diagnostic and Prognostic Health Management	Tier 2	<u>Best evidence may be required if</u>: The DHT is "first-in-class" Lack of similar platforms High consequences of inaccurate information	Initial validation: Accuracy and precision of the DHT against a well-established reference (gold) standard. Able to detect clinically relevant differences.	RCT demonstrating that use of the diagnostic DHT improves patient outcomes. If measurement of patient outcomes not feasible (e.g., long disease latency), then studies evaluating changes in clinician diagnostic impression and clinician behavior may be acceptable.
Professionally Directed Preventative and Therapeutic Health Management	Tier 3a	<u>Best evidence may be required if</u>: The DHT is "first-in-class" Lack of similar platforms High consequences of ineffective intervention	High quality observational or quasi-experimental studies with an appropriate comparator and relevant patient outcomes. Outcomes may include patient reported outcomes, engagement with the healthcare system, or clinical data.	RCT demonstrating clinical efficacy. Study may be conducted in a selected population. Surrogate outcomes and short-term follow-up may be acceptable.
	Tier 3b		RCT demonstrating clinical efficacy. Study may be conducted in a selected population. Surrogate outcomes and short-term follow-up may be acceptable.	Additional RCT demonstrating clinical effectiveness and generalizable to the patient population and health systems of interest. Demonstration of durability of treatment effect to pre-established timepoint guided by expert opinion. Captures patient-centered outcomes and treatment is compared to best or most common active comparator.

DHT: Digital health technology, RCT: Randomized Controlled Trial

<u>Safety and Effectiveness Evidence Standards, Tier 1</u>

Tier 1 is comprised of DHTs that perform self-directed health management. DHTs in this tier are products that are available for individual use without clinician involvement. Software applications that can be downloaded and are free or paid for by an individual most commonly fall into this tier. This includes personalized health information for home use (e.g., symptom monitoring or risk assessment tools) that can be delivered via the Cloud/Internet/mobile app. Wearable devices for personal use are included in this tier of evidence when they are initiated and used by an individual without direct engagement of a clinician in interpreting or acting urgently on the results. Common tier 1

DHTs include fitness trackers, apps that utilize smart watches, or simple symptom diaries that do not share information with a healthcare provider, as well as mobile apps that deliver information that is personalized to the user (e.g., ovulation tracking). Other consumer-driven self-directed health apps (e.g., virtual reality as vision therapy or apps to prevent cognitive decline) also fall under this tier.

Tier 1 DHTs are universally low risk since they are user-initiated and not integrated into clinical care, although DHTs in this tier may also measure or attempt to impact outcomes related to health (e.g., tobacco use) or behavior (e.g., hours slept, number of steps). The minimum evidence standard requires only that the DHT provides accurate information (e.g., heart rate measured from the mobile app gives accurate and reliable information compared to a gold standard). The best evidence standard for these DHTs would also include empirical data on the *perceived* behavior change or health improvement by end-users. Studies to generate this kind of data will typically follow a pre- / post- study design within users; outcomes can be either self-reported or captured objectively (e.g., physical activity/ number of steps). Quantitative valid measures are preferred for self-reported outcomes.

Safety and Effectiveness Evidence Standards, Tier 2

Evidence Tier 2 is the first of two tiers that evaluate the safety and effectiveness of DHTs whose use directly involves health professionals. Tier 2 is focused on diagnostic and prognostic DHTs, recognizing that DHTs with these functions require a separate type of evidence than those that are therapeutic. General functions of DHTs in this category are to diagnose a specific clinical condition or to provide prognostic information, often by monitoring some physiologic process, a function often called "active monitoring." This includes implants or sensors worn on the body that transmit information to aid in diagnosis or management (e.g., digital patch for noninvasive blood glucose monitoring, or an intelligent patch that can detect early signs of breast cancer and transmit the information to a lab for analysis). Tier 2 DHTs also include technology that guides treatment decisions, such as software that uses data to diagnose a condition (e.g., software that uses machine learning to rapidly diagnose a myocardial infarction on ECG) or digital algorithm that guides clinical decisions for a healthcare professional (e.g., software that predicts response to chemotherapy regimens). A key distinguishing feature of this tier from the following is that while data may influence clinician decision making through diagnosis or prognosis, it does not intervene directly in the care of the patient.

Tier 2 DHTs are considered moderate to high risk for individual patients (no longer called "end-users"), driven by the consequences of inaccurate diagnostic or prognostic information. For diagnostic DHTs at the lower bounds of moderate risk, minimum evidence only requires the evaluation of accuracy and precision against a well-established gold standard. Better evidence would include external validation of the DHT in a second, independent population with different background clinical characteristics. However, DHTs posing higher risk require greater certainty that the information provided produces improved health outcomes within reasonable risk related to false positive and false negative findings. For example, with high-risk diagnostics (e.g., cancer diagnosis), an RCT demonstrating that use of the diagnostic DHT improves patient outcomes may be required unless there is no risk of increased false positives or negatives with the new product. In such cases, or if measurement of patient outcomes is not feasible (e.g., long disease latency), then studies evaluating changes in clinician diagnostic impression and clinician treatment decisions are acceptable surrogates to capture the impact of the DHT.

Safety and Effectiveness Evidence Standards, Tier 3a and 3b

Tier 3 encompasses professionally-directed preventative and therapeutic health management interventions. These DHTs have direct clinician involvement and measurable clinical outcomes. However, this Tier (3a and 3b) has been

sub-divided to further distinguish function and risk within this broad class. Tier 3a is comprised of interventions for health behavior management and prevention, whereas Tier 3b consists of therapeutic interventions for an established clinical diagnosis. There are some cases that may seem fitting for both tiers (e.g., an app that helps with smoking cessation may be very similar to an app that is a treatment for a nicotine use disorder), but key features can help determine the appropriate evidentiary standards. To help distinguish which evidence tier is most fitting, one should determine the intended use (3a – modifying a behavior or prevention, versus 3b – treatment) and potential risk to patients (3a – low to moderate, versus 3b – moderate to high).

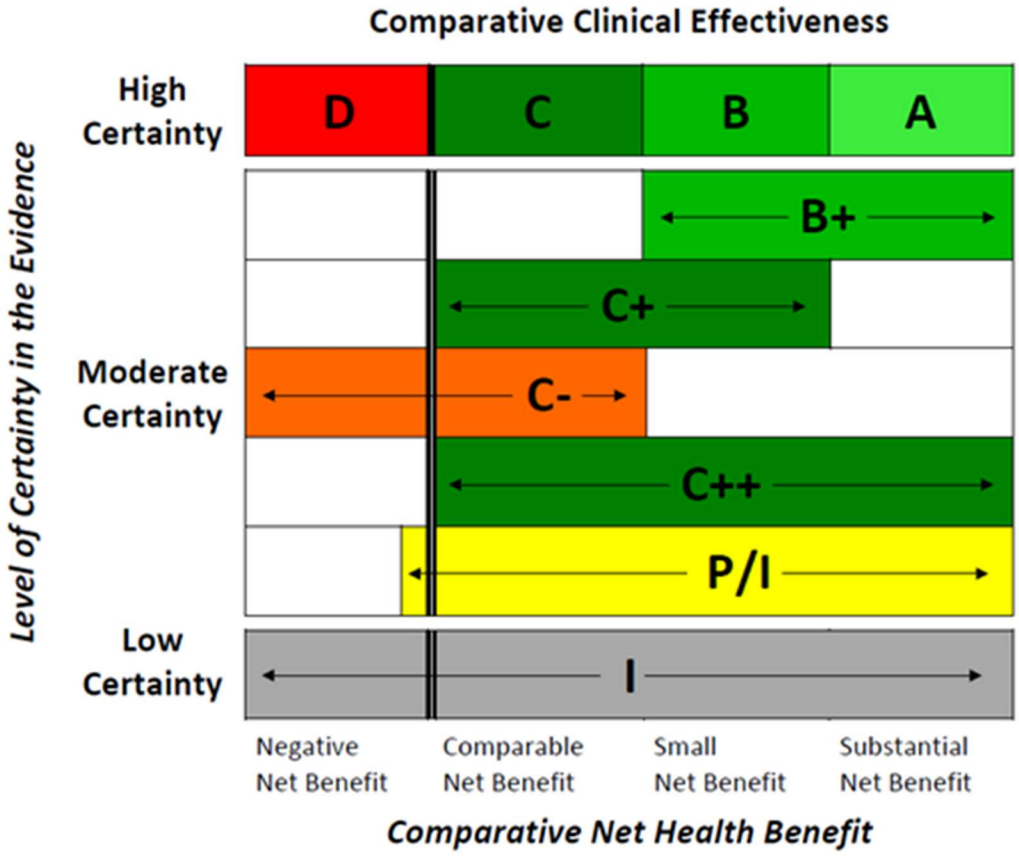
Tier 3a interventions are for preventative health or health behavior management under clinician supervision. Clinician involvement is key to distinguishing Tier 3a/3b apps from the vast number of self-directed health behavior apps available for direct download (Tier 1 DHTs). Interventions in this category are of low to moderate risk for individual patients. Those Tier 3a DHTs that are low risk should, as a minimum, have evidence from high quality observational or quasi-experimental studies with an appropriate comparator and assessment of clinically relevant patient outcomes. Tier 3a DHTs that are moderate risk should have evidence from an RCT demonstrating clinical efficacy. It is acceptable for RCTs to be conducted in a selected population (e.g., healthier individuals); surrogate outcomes and short-term follow-up may also be acceptable.

Tier 3b interventions are therapeutic interventions for clinically diagnosed conditions and they are considered moderate to high risk. DHTs in this category may undergo regulatory approval as a prescription digital therapeutic, but that is not required to be considered a Tier 3b intervention. Tier 3b interventions are therapeutic (e.g., app that delivers CBT for opioid use disorder) or directly guide a medical or surgical intervention (e.g., software that creates 3D reconstruction images and determines location for optimal needle placement for a lung biopsy). The minimum evidence requirement for moderate risk Tier 3b interventions is a well-conducted RCT demonstrating clinical efficacy. As for high-risk Tier 3a interventions, the study may be conducted in a selected population using surrogate outcomes, and short-term follow-up may be acceptable. High-risk Tier 3b interventions require at least one additional RCT that evaluates the clinical effectiveness of the intervention in patient populations and health systems of interest. RCTs for all Tier 3b DHTs should capture patient-important outcomes, compare the new DHT to a reasonable active comparator (e.g., well-managed standard of care), and demonstrate durability of treatment effect to a pre-established timepoint guided by expert opinion.

Evidence Ratings for Safety and Effectiveness

The general approach to the analysis of evidence on the safety and effectiveness of DHTs follows the precepts of ICER's Value Assessment Framework for other interventions. Evidence ratings on safety and effectiveness will follow the ICER Evidence Rating Matrix™. The ICER matrix was developed and has evolved over time with input from payers, purchasers, patient groups, and life science companies. Following a similar format used by the US Preventive Services Taskforce, the rating arises from a joint judgment of the level of confidence provided by the body of evidence and the magnitude of the net health benefit — the overall balance between benefits and harms.

Figure 2. The ICER Evidence Rating Matrix™



- ***A = "Superior" –*** *High certainty of a substantial (moderate-large) net health benefit*
- ***B = "Incremental" –*** *High certainty of a small net health benefit*
  C = "Comparable"- *High certainty of a comparable net health benefit*
- ***D= "Negative"-*** *High certainty of an inferior net health benefit*
- ***B+= "Incremental or Better" –*** *Moderate certainty of a small or substantial net health benefit, with high certainty of at least a small net health benefit*
- ***C+ = "Comparable or Incremental" –*** *Moderate certainty of a comparable or small net health benefit, with high certainty of at least a comparable net health benefit*
- ***C- = "Comparable or Inferior"*** *– Moderate certainty that the net health benefit is either comparable or inferior, with high certainty of at best a comparable net health benefit*
- ***C++ = "Comparable or Better" –*** *Moderate certainty of a comparable, small, or substantial net health benefit, with high certainty of at least a comparable net health benefit*
- ***P/I = "Promising but Inconclusive"*** *– Moderate certainty of a small or substantial net health benefit, small likelihood of a negative net health benefit*
- ***I = "Insufficient" –*** *Any situation in which the level of certainty n the evidence is low*

Health Equity: Accessibility

Like user experience, health equity underpins clinical effectiveness through its impact on DHT uptake, utilization, and subsequent health outcomes. Accessibility and inclusivity are the lens through which this framework examines whether a DHT is culturally and linguistically appropriate, has a low barrier to entry for digital literacy, instills or exacerbates implicit biases, and is adaptable to meet the usability needs of health disparity populations. Per the National Institutes of Health, these populations include racial and ethnic minority groups, socioeconomically

disadvantaged populations, underserved rural communities, and sexual and gender minority groups. It is here that the distinction between "access" and "accessibility" is made.

In the case of apps and similar DHTs, there are several well-established disability accessibility standards such as the Web Content Accessibility Guidelines (WCAG). These standards dictate that for a product to be accessible it must be perceivable (related to the senses, most often sight or hearing), operable, and understandable (to either the person or an assistive agent or device). If any of these do not hold, then users with disabilities or impairments will not be able to use the DHT. Assessment of accessibility will also include noting the availability of the product in multiple languages (when applicable) and the reading level of the language the end user must navigate. The CDC has developed the CDC Clear Communication Index, a research-based tool that can assess how accessible public communication materials are based on the health literacy of the public. The tool assesses the material and assigns a 0 to 100 score where 90 is passing, but 100 is ideal to be accessible.

Assessment of accessibility will consider established accessibility standards but will be qualitative across the DHT(s) being assessed in a report.

Health Equity: Access & Distribution

The ICER-Peterson framework also includes evaluation related to the equity implications of the DHT for access across different patient subpopulations and distributional areas, such as whether a DHT can be administered effectively in rural areas, to individuals with limited access to broadband internet and other technologies, or to socioeconomically disadvantaged populations.

DHTs may also affect health equity if they target conditions or broader socioeconomic determinants of health that have disproportionate burden for certain populations. DHTs may have the ability to increase access through multiple routes, including making care more available without requiring access to specialist clinicians or other sites of care. Successful targeting of these conditions or factors may lead to reductions in disparities in health outcomes that have plagued the US health care system. However, some DHTs will have the potential to widen disparities. DHTs that require advanced computer or cell phone technology or user expertise may only add to the digital "gap" between more affluent and lower-income individuals.

Ideally, evidence on access and distribution will be available to evaluate whether a DHT is likely to affect health equity. This framework intends to leverage work published by ICER in March 2023 on advancing health technology assessment methods by improving considerations of health equity across different steps of an HTA review. Developers should seek to design DHTs to meet the needs of diverse populations and test DHTs among as broad a spectrum of patients, clinicians, and settings as feasible. Assessment of the impact on access and distributional effects of DHTs will be qualitative.

Economic Impact

If DHTs are to be paid for by plan sponsors, provider groups, or insurers, they should have a robust dossier of evidence demonstrating the economic impact of the product across the breadth of the health system. Unlike new drugs, which are viewed as producing value through health gains at a price that almost always adds cost to the health system, payers in the US market assume that DHTs should produce a combination of improved care outcomes and reduced overall costs, or provide substantially improved access with costs lower than other options. This different paradigm for value means that the methods of assessing the economic impact of DHTs as an element

of its “value” will emphasize budget impact analysis rather than long-term cost-effectiveness analysis. In some cases, the economic impact of a DHT may also generate other benefits, such as improvements in productivity or reductions in family and caregiver time and expenses. These benefits will be captured in the economic impact assessment but will not be included in the primary budget impact analysis.

The budget impact analysis of a DHT estimates its incremental costs (positive or negative) versus its comparator in a stated population, from a stated perspective, and over a specified time horizon. Under the ICER-Peterson DHT Framework, DHTs that add costs while expanding access significantly to effective care will have a budget impact scenario analysis comparing the added costs to the estimated costs that would otherwise be necessary to expand access to the same degree by augmenting the current standard of care.

The outputs of the budget impact analysis should be presented in a standard format to facilitate comparison across DHTs. Table 4 on the following page illustrates the framework for use in the budget impact analyses of DHTs. Sections following this table describe in detail the considerations regarding the perspective and time horizon for DHT budget impact analyses.

Table 4. Standard DHT budget impact analysis framework and outputs	
Budget Impact Feature	***Comments***
Budget impact outputs • Incremental cost (e.g. with vs. without DHT) per end user/patient adhering to DHT • Incremental cost per 1,000 users reflecting adherence • Incremental cost per user per month • Scenario at threshold price to achieve different levels of cost savings • Scenario(s) within proposed outcomes-based agreement	• Incremental costs with vs. without DHT shown with impact on subcategories of cost: a. Intervention costs b. Impact on standard of care costs c. Costs related to changes in access to care • Outputs generated at given or estimated price
Perspective • Individual end user/patient • Third-party payer/plan sponsor	• Typical user cost subdomains: a. user time costs b. transportation costs c. other expenditures due to changes in health management costs • Typical payer cost subdomains: d. medical e. diagnostic f. pharmacy costs g. relevant provider productivity costs
Time Horizon • Year 1 with scenario analyses for implementation costs that may differ across practice settings • Year 2-5 if clear rationale for extended horizon	• Framework allows flexibility in time horizon to fit the DHT and its environment alongside cost impacts.

Analysis Outputs

To present a broad view of the economic impact of DHTs the output of budget impact analysis will be framed in three different ways. First, the incremental costs will be calculated at the level of the individual end user or patient adhering to the DHT. Although the more relevant economic impact for most purposes will be calculated on the entire population initially assigned to use a DHT, or the population to which a DHT is made available, it may still be of interest to understand the cost impact for those users/patients who use the DHT as intended with full adherence.

For payers and plan sponsors it will be most useful to understand the budget impact at the population level, either through a standardized cost impact per 1,000 users/patients prescribed or assigned the DHT, or from the impact on the overall health cost as expressed as a change in the per user per month (PUPM) expense. When made feasible by the size of the patient/enrollee population studied, subgroup analyses may be performed to evaluate whether the budget impact varies for key subpopulations. For DHTs that will be assigned to the entire plan population, the PUPM will equal the per member per month (PMPM) measure. The PMPM impact is particularly useful to payers because it is used to compare the effect of multiple different types of care interventions on health spending, and it is a common metric used in describing opportunities for health care savings to employers and other plan sponsors. Many DHT developers will seek adoption and coverage for products based on their own analyses of the potential

PUPM or PMPM savings, so comparing these claims to the results of an independent budget impact analysis will be an important part of the assessment of economic impact.

Whenever possible, budget impact analyses will be performed using prices already established in the market. Prior to market adoption, placeholder prices will be used based on pricing suggested by the developer or by analyst predictions. Price benchmarks will also be calculated to meet specific cost savings targets, a method discussed in greater detail below.

Perspective

Budget impact analyses will be performed from two different perspectives. The first is the individual "DHT user" perspective. Although professionally directed DHTs may have limited or no economic impact on out-of-pocket costs for the individual whose care is informed or delivered by a DHT, for some of these DHTs there may be an important impact on the number of clinician visits, medications used, etc. Assessing the economic impact on individuals is challenging due to varying benefit designs. In these cases, the assessment may either use the most common cost sharing structure or create a range based on different benefit designs. Typical cost domains that will be measured under the DHT user perspective in addition to the cost of the DHT and will include: user time costs, transportation costs, and other out of pocket expenditures due to changes in health management costs.

The second and likely most emphasized perspective examined in each budget impact analysis will be that of the payer, either health plans/PBMs or plan sponsors (e.g., employers). Under this perspective all health care-related costs (e.g., medical, diagnostic, and pharmacy costs) will be included over the specified time horizon for the DHT and its comparator. DHTs may have direct or indirect effects on health care practitioner productivity and these effects will be modeled to the extent that such costs are relevant to the stated payer/plan sponsor perspective. If they are not relevant, health care practitioner productivity will be included in the assessment outside the quantified budget impact. When there are major differences in the cost structure for commercial payers versus public payers such as Medicaid or Medicare, alternative payer cost structures may be included in additional scenario analyses.

Time Horizon

Budget impact will be calculated over a horizon of one year with scenarios performed as needed to capture differential implementation costs across different types of health systems (if any). Scenarios will also be calculated over longer time horizons of 2-5 years should there be specific circumstances suggesting that a significant proportion of cost offsets from DHT implementation will only accrue over this longer time frame.

Implementation costs included in the analysis will include initial costs for software integration, provider and/or end user education, and effects on provider productivity should implementation require additional time or effort. If there are no or very limited implementation costs, this one-year horizon can provide a view of the budget impact for a DHT in a "steady state." An additional scenario with a time horizon up to five years can be performed if there is a clear rationale that cost offsets from the DHT grow over time, perhaps as patients accrue longer-term clinical benefits suggested by short-term changes in surrogate outcomes such as blood pressure or hemoglobin A1C. For all time horizons, outputs will be calculated based on the experience of one cohort from inception to avoid changes in outputs based on unpredictable variation in DHT uptake in the population. Multiple yearly cohorts of DHT users and its impact on the budget impact findings may be explored in uncertainty analyses, given sufficient rationale.

Threshold Pricing for Cost Savings

As noted earlier, budget impact analyses will be performed using market-based prices or estimates provided by DHT developers or analysts. The threshold price at which a DHT produces cost savings to the payer will be calculated. Based on initial discussions with payers and plan sponsors on relevant price points for consideration of DHT adoption, it is clear that some decision-makers use a return-on-investment target in considering reimbursement levels, whereas others seek to apply a per-user-per-month or overall cost savings framework. An overall cost savings approach may offer greater advantages as a starting point for pricing considerations. There is no single target for cost savings that will reflect a "reasonable" or "fair" price in every situation, but discussions with payers have suggested that a good starting point would be a price estimated to achieve an overall 15% cost savings versus usual care for the patient or pool of patients using the DHT. This level of estimated cost savings reflects the inherent uncertainty regarding the real-world impact of many DHTs and the sunk time and resource costs of implementing them on a widespread basis. A threshold price at which a DHT under review would produce 15% cost savings will then be calculated. For DHTs targeting patients with very high annual costs under standard of care, a 15% cost savings target may not be deemed possible, and in such cases a threshold price to produce a return-on-investment of 3:1 will be calculated, a target also suggested by payers as common in the market as a threshold representing good value. Whether a percent savings or return-on-investment is calculated, it should be noted that the threshold price may vary across different time horizons based on the timing of any cost offsets following implementation of the DHT.

As acknowledged earlier, a budget impact perspective is a different approach than taken for new drugs which generally add costs to the health care system, and where threshold price estimates are generated using an opportunity cost threshold (in the US context) of $100,000 to $150,000 per added quality adjusted life year or equal value life year gained.

If there is no positive DHT price that will achieve 15% cost savings and/or a return-on-investment of 3:1, a threshold price that achieves cost neutrality will be estimated. For assessments that include more than one DHT, multiple pairwise comparisons are possible. Rationale will be provided to justify the relevant pairwise comparisons for pricing thresholds within the scoping phase of the assessment.

It is important to note that variations across clinical and health system contexts, and differing levels of uncertainty regarding the underlying evidence on clinical and budgetary impact, will affect the view of decision-makers on which level of overall cost savings is most relevant, and on how to apply calculated threshold prices in negotiations with DHT developers.

Payment Agreement Scenarios

Some DHTs will be introduced with a variety of pricing and payment options for payers. Possible options might include an initial price for an entire patient population with rebates or milestone payments based on uptake and long-term use. Other agreements may specify per-user payment fees that are reduced once a certain number of users is reached. Assessors will seek information about potential pricing and payment agreements from developers and payers and will and include these as scenarios in the budget impact analyses. If the developer is launching their DHT with one dominant pricing/payment arrangement, it will be considered as the base case for calculations of threshold pricing.

Additional Scenario Analyses

Given that there are frequently limited data on many of the resource utilization or cost inputs needed to generate a budget impact analysis, structural scenario analyses are viewed as more valuable in assessing uncertainty than

the one-way sensitivity and probabilistic sensitivity analyses that are commonly conducted in cost-effectiveness analyses. Therefore, additional scenario analyses will be listed in a model analysis plan and carried out with comparable outputs as the base case findings. Examples of scenario analyses include: different risk profiles of end users, changes to other resource use, changes to the time horizon when there is uncertainty in the potential impacts of downstream cost savings, changes to the upfront costs to adopt the DHT, and changes around whether certain costs are time varying or not. These scenario analyses will give a more comprehensive understanding of the plausible range of budget impact while also providing insights into model components most likely to impact the findings.

Summary Ratings and Action Recommendations

Decision-makers should be presented with an analysis allowing them to understand the strengths and limitations of the evidence for each of the components of the value framework described in this document. Additionally, several different ways to present overall "summary" ratings have been considered for the broader categories of clinical and economic impact. The difficulty with summary ratings is that they must somehow integrate both the magnitude of the benefit or disadvantage of a DHT relative to some comparator with the strength of the evidence supporting that difference. This "dual axis" problem is the reason that the safety and effectiveness rating will be presented as a single grade that represents a combination of these two factors. This approach, however, does not fit the way that evidence and perspectives will be assessed for judgments regarding user experience or health equity.

Therefore, decision-makers will likely find it more useful to have a final recommendation related to adoption rather than some kind of composite overall rating or a formal "heat map" across all the dimensions in the value framework. The final summary of the assessment will therefore include one of the following recommended actions based on the evidence:

1. Evidence inadequate to support broad field testing
2. Evidence adequate to support field testing in broader populations
3. Evidence adequate to support wide adoption

Each of these recommendations will be followed by text providing the rationale for the recommendation, highlighting key uncertainties remaining in the evidence base. When relevant, suggestions for specific research studies to address these limitations will be given. The goal will be to provide decision-makers with a summary recommendation that contains enough detail so that each decision-maker will be able to parse whether the recommendation is a good fit for their organization given its unique needs, characteristics, and relationship with the DHT developer.

Ultimately, the framing of recommended actions will evolve as purchasers and payers of DHTs, and DHT developers themselves, provide further input on how such overall judgments can best provide actionable information without obscuring the important details and nuance within each of the various domains of this value assessment framework.

References

Administration, U. F. (2020, September 22). *U.S. Food and Drug Administration*. Retrieved February 28, 2023, from What is Digital Health?: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health

Administration, U. F. (2021, November). *Content of Premarket Submissions for Device Software Functions*. Retrieved February 28, 2023, from U.S. Food and Drug Administration Guidance Documents : https://www.fda.gov/regulatory-information/search-fda-guidance-documents/content-premarket-submissions-device-software-functions

Administration, U. F. (2022, December 8). *Medical Devices*. Retrieved February 28, 2023, from Digital Health Center of Excellence: https://www.fda.gov/medical-devices/digital-health-center-excellence

Andalucía, A. d. (2012). *Safety and Quality Strategy in Mobile Health Apps*. Retrieved February 28, 2023, from Complete list of recommendations on design, use and assessment of health Apps: http://www.calidadappsalud.com/en/listado-completo-recomendaciones-app-salud/

Care, A. G. (2022, August 17). *How we regulate manufacturing*. Retrieved February 28 , 2023, from Regulation of software based medical devices. Software based medical devices are medical devices that incorporate software or are software.: https://www.tga.gov.au/how-we-regulate/manufacturing/medical-devices/manufacturer-guidance-specific-types-medical-devices/regulation-software-based-medical-devices

Devices, F. I. (2020, August 7). *BfArM*. (F. I. (BfArM), Producer) Retrieved February 28, 2023, from The Fast-Track Process for Digital Health Applications (DiGA) according to Section 139e SGB V A Guide for Manufacturers, Service Providers and Users: https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.pdf?__blob=publicationFile

Division, E. M. (2020, June 1). *European Medicines Agency*. Retrieved February 28, 2023, from Questions and answers: Qualification of digital technology-based methodologies to support approval of medicinal products: https://www.ema.europa.eu/en/documents/other/questions-answers-qualification-digital-technology-based-methodologies-support-approval-medicinal_en.pdf

Excellence, N. I. (2023). *Our Programmes*. Retrieved February 28, 2023, from Evidence standards framework (ESF) for digital health technologies: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies

Group, I. S. (2014, September 18). *International Medical Device Regulators Forum*. Retrieved February 28, 2023, from Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf

Hajesmaeel-Gohari, S. K. (2022, January 27). The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. *BMC Med Inform Decis Mak*. doi:https://doi.org/10.1186/s12911-022-01764-2

U.S. Department of Health and Human Services Food and Drug Administration, C. f. (2005, May 11). *Guidance for Industry and FDA Staff*. Retrieved February 28, 2023, from Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices: https://www.fda.gov/media/73065/download

U.S. Food and Drug Administration. (2021, January). *Software as a Medical Device (SaMD)*. Retrieved February 28, 2023, from Artificial Intelligence and Machine Learning in Software as a Medical Device: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

Usability.gov. (n.d.). *How to and Tools*. Retrieved February 28, 2023, from System Usability Scale (SUS): https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

WAI, W. C. (2023, January 25). *Standards/Guidelines*. Retrieved February 28, 2023, from WCAG 2 Overview: https://www.w3.org/WAI/standards-guidelines/wcag/