Original Research Article

Assessing the Value of Provider-Facing Digital Health Technologies Used in Chronic Disease Management: Toward a Value Framework Based on Multistakeholder Perceptions

Caitlin Main⬤, Madeleine Haig, Danitza Chavez, and Panos Kanavos

Objectives. Hardly any value frameworks exist that are focused on provider-facing digital health technologies (DHTs) for managing chronic disease with diverse stakeholder participation in their creation. Our study aimed to 1) understanding different stakeholder opinions on where value lies in provider-facing technologies and 2) create a comprehensive value assessment framework for DHT assessment. **Methods.** Mixed-methods comprising both primary and secondary evidence were used. A scoping review enabled a greater understanding of the evidence base and generated the initial indicators. Thirty-four indicators were proposed within 6 value domains: health inequalities (3), data rights and governance (6), technical and security characteristics (6), clinical characteristics (7), economic characteristics (9), and user preferences (3). Subsequently, a 3-round Web-Delphi was conducted to rate the indicators' importance in the context of technology assessment and determine whether there was consensus. **Results.** The framework was adapted to 45 indicators based on participant contributions in round 1 and delivered 16 stable indicators with consensus after rounds 2 and 3. Twenty-nine indicators showed instability and/or dissensus, particularly the data rights domain, in which all 5 indicators were unstable, showcasing the novelty of the concept of data rights. Significant instability between *important* and *very important* ratings was present within stakeholder groups, particularly clinicians and policy experts, indicating they were unsure how different aspects should be valued. **Conclusions.** Our study provides a comprehensive value assessment framework for assessing provider-facing DHTs incorporating diverse stakeholder perspectives. Instability for specific indicators was expected due to the novelty of data and analytics integration in health technologies and their assessment. Further work is needed to ensure that, across all types of stakeholders, there is a clear understanding of the potential impacts of provider-facing DHTs.

Corresponding Author
Caitlin Main, Department of Health Policy and Medical Technology Research Group, LSE Health, London School of Economics and Political Science, Cowdray House, Houghton Street, London, WC2A 2AE, UK; (C.a.main@lse.ac.uk).

Highlights

- Current health technology assessment (HTA) methods may not be well suited for evaluating digital health technologies (DHTs) because of their complexity and wide-ranging impact on the health system.
- This article adds to the literature by exploring a wide range of stakeholder opinions on the value of provider-facing DHTs, creating a holistic value framework for these technologies, and highlighting areas in which further discussions are needed to align stakeholders on DHTs' value attributes.
- A Web-based Delphi co-creation approach was used involving key stakeholders from throughout the digital health space to generate a widely applicable value framework for assessing provider-facing DHTs. The stakeholders include patients, health care professionals, supply-side actors, decision makers, and academia from the United States, United Kingdom, and Germany.
- High levels of instability among stakeholders and value domains are demonstrated, indicating the novelty of assessing provider-facing DHTs and their impact on the health system.

Value frameworks have been increasing in prevalence throughout the health sector as a technique for evaluating different health technologies to inform resource allocation.[1–3] Their construction typically involves incorporating a range of crucial stakeholder opinions to create a holistic understanding of what value means in regard to specific medical technologies.[4] It is believed that their proliferation is partly a result of the apparent disconnect between price and value, increasingly complex technologies, and greater empowerment of clinician and patient opinions of potential technologies.[1] Our health systems are increasingly tasked with managing chronic diseases. This in part has led to the recent focus on value-based health care, and there is an increasing emphasis on patient-centered care, and what value means to them; thus, patient voices are thought to be one of the most important in creating value frameworks.[5] Digital health technologies (DHTs), with their significant variation in risk level and functionality, fast-paced innovation, short-life cycles, and limited evidence-creation capabilities, are seemingly the perfect candidate for a widely applicable value framework.[6,7]

Evaluating DHTs for chronic disease management must involve consideration of value domains beyond traditional economic and clinical impact due to the additional risks imposed by their shared ability to generate, store, process, and/or transmit data.[8] These risks include data privacy, security, governance, transparency, and bias and have not previously been an issue when assessing health technologies such as medical devices and pharmaceuticals.[9] This technological shift raises a critical question: what new aspects of value arise from introducing DHTs in the health sector? Does the introduction of widespread big data collection and analysis require a unique, and possibly specialized, approach to evaluation, or can evaluation methods for traditional medical technologies simply be adapted to suit digital innovations?

Although the health sector has been notably sluggish in its adoption of and attention to digital technologies compared with other sectors, DHTs have risen sharply in prevalence worldwide. They are considered crucial for health systems to reach their key objectives of improving

Department of Health Policy and Medical Technology Research Group, LSE Health, London School of Economics and Political Science, London, UK (CM, MH, DC, PK). The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this paper was provided by a contract with Sanofi S.A. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

effectiveness, equity, efficiency, and quality of care.[10] Their potential is significant due to wide-ranging functionality and applicability to several different actors, such as patients, health care professionals (HCPs), industry, and health systems as a whole.[11] As a result of this significant technical and functional variation, decision makers and health technology assessment (HTA) agencies are finding it difficult to accurately assess the value of these technologies.[7]

It is difficult to determine a methodology for assessing DHTs that have a significant and far-reaching impact on the health system. Provider-facing DHTs are often highly complex and involve multiple components, some of which evolve over time. Current HTA methods primarily exist to evaluate medical technologies used to treat individual patients. These methods rely primarily on incremental cost-effectiveness ratios, which compare a new technology's costs and clinical effectiveness with an existing course of treatment. For some DHTs, such as artificial intelligence (AI)–assisted diagnostic imaging technologies, this comparative approach is partially appropriate and allows for the quantification of the value of digital innovation. However, many DHTs also offer potential impact far beyond incremental improvements in treatment pathways for individual patients. Many provider-facing solutions have the potential to produce high-impact health data analytics due to their comprehensive and sometimes constant data collection features. For example, AI-based clinical decision-making support tools are constantly learning and evolving, so they not only offer value in the form of today's improved diagnostic accuracy but also in the form of future diagnostic improvements, learnings across the population, and potential systemwide monitoring resource for disease progression. The continuous learning and adaptability of these technologies have direct and indirect impacts on patient outcomes and health care delivery, making it difficult to assess their long-term effect on the health system.

Many DHTs have the potential to produce real-world evidence (RWE) that eventually could revolutionize the way health systems approach disease detection and diagnosis. Currently, RWE sources are primarily electronic health records, insurance claims data, patient registries, and pharmacy databases.[12] DHTs offer an opportunity for RWE directly from the digital technology, which could include data with undeniable provenance and high accuracy in a nonclinical setting. However, this wealth of data also raises data governance issues regarding which actors have the right to access and control sensitive information as it is shared across stakeholders. For example, to what level of detail should information from a patient's closed-loop insulin delivery system be available for analysis in population health management systems? Because of these privacy, security, and ethics concerns, it is important to make use of an advanced digital infrastructure that uses technology such as federated machine learning, which does not require centralized data warehouse facilities,[13] and zero-knowledge proofs, which offer information verification without data transfer,[14] for example. The ability to contribute to RWE is a technology value proposition that may not be comprehensively accounted for under current HTA methodologies. Provider-facing DHTs often produce valuable effects that can be identified across health systems, but these aspects of value may be difficult to provide strong evidentiary support for and quantify. Where widespread data collection is involved in DHTs, data management complications arise in which the ownership of and responsibility for the data becomes blurry. This issue highlights the need for reliable health system data infrastructure, which is neither the responsibility of a single stakeholder group nor solvable through HTA incentives alone.

This article aims to understand better where key stakeholders (not just decision makers) perceive value in regulated provider-facing DHTs and, ultimately, create an agnostic value framework to assist with their assessment. We have recently published an alternate article[15] focusing on assessing patient-facing technologies, which present different benefits and challenges for HTA. In comparison, provider-facing technologies offer solutions with a broader impact than patient-facing technologies because they often involve systemwide integration into care delivery pathways as opposed to personalized solutions for individuals.

This article adds to the literature in 3 ways: first, it explores a wide range of stakeholder opinions of value aspects of provider-facing technologies; second, it explores a holistic value framework of these system-facing DHTs that incorporates a significant number of different types of technologies; and third, it highlights where stakeholders disagree or are unsure about the value of these technologies, ultimately allowing policy and decision makers to understand where conversations need to be focussed.

The following section outlines the methods of our study and is followed by the results of the literature review and Delphi exercise. A discussion of key issues discovered in the exercise occurs before the conclusion.

Methods

This study, carried out between September 2021 and April 2022, had 3 defined stages using a literature review

Table 1 Stages of the Project

	Components	Objectives	Methods	Tasks
First stage	Scoping literature review	Identify value frameworks (VFs) oriented toward digital health technologies (DHTs) Identify value domains and indicators from VFs	Qualitative	Overview of DHT value frameworks through a targeted scoping review focusing on the 3 study countries
	Initial list of indicators	Create an initial value framework	Qualitative	Researcher consensus meeting to create indicators and a brief "more information" paragraph for each indicator
	Indicator categorization	Categorize indicators chosen for value framework	Qualitative	Use scoping review to identify "value domains" to categorize chosen indicators
	Stakeholder identification	Identify and invite stakeholders	Qualitative	Identify and invite stakeholders from the London School of Economics network in the 5 stakeholder groups (patients, health care providers, supply-side actors, decision makers, and "influencers") in 3 study countries (United Kingdom, United States, and Germany)
Second stage	Stakeholder proposed indicators	Identify new indicators from Delphi R1 Thematically analyze proposed indicators to identify ideas and themes not previously incorporated in VF	Qualitative/ quantitative	Delphi exercise R1—stakeholders asked to comment on existing indicators or propose their own Thematic analysis using Nvivo Software so identify common themes in comments and proposed indicators
	Updated VF	Update initial proposed VF with stakeholder proposed themes	Qualitative	Researchers consensus meeting using thematic analysis data to update VF
Third stage	Delphi exercise R2 and R3	Consensus-building exercise to understand which indicators have stable consensus	Quantitative	Delphi exercise R2—stakeholders asked to rate each indicator on an "importance" Likert scale Delphi exercise R3—stakeholders shown their own response in R2 and the groups responses and given opportunity to change their answer
	Data analysis	Analysis of respondent data within and between groups to determine respondent stability and consensus	Quantitative	Likert scale data analyzed for each indicator; consensus and stability calculated within and between stakeholder and country groups

and a Web-based Delphi exercise. The stages are as follows: 1) stakeholder identification and targeted literature review to propose an initial framework, (2) framework revision through qualitative Delphi round 1 (R1), and 3) quantitative Delphi rounds 2 (R2) and 3 (R3) for stakeholder evaluation of proposed indicators and statistical analyses (summarized in Table 1).

Decision Context

The United Kingdom (UK), United States (US), and Germany were chosen as study countries. This resulted from their recent developments regarding innovative policies and regulations for DHTs and their representation of different health system archetypes of health system financing—taxation, social insurance, and private health

insurance, respectively. Germany launched the Digital Health Care Act in 2019 tying statutory reimbursement policies to their digital health value framework. It enabled HCPs to prescribe digital applications to patients that complied with particular standards for data protection, information security, quality, and safety as well as other factors.[16] However, Germany lacks a clear pathway for system-facing technologies. Comparatively, the United Kingdom and United States do not have a national reimbursement framework for DHTs, but they have made significant efforts to enable their assessment. The United Kingdom's methodological advancements include the NICE evidence standards framework and the Digital Technology Assessment Criteria for health and social care. Alternatively, the United States has created the Digital Health Center of Excellence within the Food and Drug Administration, which aims to improve digital health policies and regulatory approaches.

Our focus was purely on provider-facing DHTs, users of which include HCPs and imaging technicians who are considered actors of health providers and the health system. These DHTs may not immediately affect a patient's daily behavior, and a patient may not be aware the technology is being used, even though their data are being collected (see Appendix Table 1). An example is AI-assisted cardiac magnetic resonance imaging software, which assists HCPs in making cardiac diagnoses. This technology can provide value to the system by increasing the prevalence of correct diagnoses and overall efficiency by reducing the diagnosis time. This downstream value must also be accounted for during assessment.

First Stage: Targeted Literature Review, Initial Framework Proposal, and Stakeholder Identification

Targeted literature review. A scoping review was completed to better understand the existing value frameworks and determine gaps in the literature and knowledge base regarding the value frameworks of DHTs for chronic conditions (see Appendix Section 2.1 and Appendix Table 2 and Appendix Figure 1 for more information regarding the literature review).

Value domains of DHTs identified in the literature include economic, clinical, and technical characteristics; user preferences; safety of the technology; and regulatory compliance (see Appendix Table 5 for more information on domains identified in the peer-reviewed and gray literature). An initial framework was proposed, adapting the value domains and indicators found in the literature. This starting framework consisted of 34 value criteria in 6 domains: 1) health inequalities, 2) data rights and governance, 3) technical and security, 4) economic

characteristics, 5) clinical characteristics, and 6) user preferences (see Appendix Table 6).

Value framework proposal through indicator identification and categorization. In this article, the term *indicator* refers to a particular characteristic considered when assessing new technologies. We define *value domain* as a category of technology assessment within which indicators are nested. The targeted literature review identified common indicators and key value domains. The research team had a consensus meeting to extract and consolidate overlapping relevant indicators and value domains to create an initial value framework. Moreover, "more information" paragraphs were created for each indicator to ensure all participants had a strong and similar understanding of each indicator.

Stakeholder identification and invitation. Five key stakeholder groups were identified and included in creating the value framework through participation in a Delphi exercise. These groups were patients (including patient advocates) and caregivers, HCPs with experience using DHTs, supply-side stakeholders (those involved in the conception/production of DHTs), decision makers (pricing and budgetary decisions for DHT reimbursement), and, lastly "influencers" (academics and policy experts). Appendix Section 2.2 and Appendix Table 3 discuss eligibility criteria.

Participants were primarily identified through the London School of Economics (LSE) networks; however, assumptions around individuals' willingness to participate were made at the beginning of the project that did not hold throughout the recruitment period. We found that Americans and Germans from the patient, HCP, and decision-maker stakeholder groups were not easily agreeable to participate in the study; the strategy was thus changed, and professional recruiters were hired to ensure equal participation from the different groups. The research team decided the importance of equal representation outweighed potential biases that recruiters may have introduced and constitutes a limitation. Please see Appendix section 2.2 for more information regarding recruitment.

Stage 2: Value Framework Update through Stakeholder Input

The Delphi approach was used to elicit the preferences and value concerns of different stakeholder groups. Delphi studies are used to measure consensus and dissensus of participants regarding a particular topic throughout

several "rounds" of the exercise.[17] When using this method, before determining consensus levels, it is important to 1) have an open qualitative round for stakeholders to submit their opinions and ideas; and 2) establish the stability of participant responses between rounds because if a participant's opinion is actively changing, levels of agreement can change in response.[17,18] We used the online platform Welphi[19] to perform our 3-round Web-based Delphi exercise.

Stakeholder proposed indicators. In R1 of the Delphi exercise, stakeholders were shown our initial proposed value framework consisting of 34 indicators in 6 domains and were asked to comment on and propose new indicators. NVivo software[20] was used to thematically analyze proposed indicators to incorporate in the updated value framework. The themes were incorporated either by creating new indicators or assimilation into previously proposed indicators.

Stage 3: Quantitative Analysis of Framework

Likert scale data were collected through R2 and R3 of the Delphi exercise and were analyzed according to participant stakeholder and country groups.

Delphi rounds 2 and 3. In R2, participants were presented with the updated framework and asked to rate each indicator on a 5-point "importance" Likert scale (1–5, with 1 being *not at all important* and 5 being *very important*) according to the decision context proposed. In R3, participants were shown how they responded to each indicator and the distribution of responses across the entire group. They then could change their answer or keep it the same.

Statistical analyses. Several statistical tests were completed using STATA 16.1 software[21] to determine stakeholder and country group differences. First, the interrater agreement (IRA) of participants within each group was determined through the Kappa statistic and Gwet's agreement coefficient with linear weights, using a benchmark scale to assess the levels of agreement.[22] This enabled us to compare groups with sufficient internal agreement by establishing whether individuals within each stakeholder and country group had a substantial likelihood of making the same judgment independently.

Second, Wilcoxon's test was used to assess the stability of group responses for each indicator between R2 and R3, showing whether respondents were confident in their judgments or were actively changing their minds

and considering new viewpoints. Stability was considered a prerequisite to consensus measurement, as unstable indicators warrant further exploration in additional Delphi rounds.[17,18,23] An indicator is considered unstable if ≥ 1 subgroup is unstable according to the Wilcoxon test.

Third, consensus or dissensus of each indicator for the entire group was determined by the interquartile range (IQR) and median responses for each indicator. Consensus was defined if an indicator was stable and IQR was ≤ 1. Subgroup analysis using the Kruskall-Wallis H-test was conducted to determine whether there were significant disagreements between groups for each indicator, and Dunn's test was used to establish which groups disagreed. An indicator could have consensus based on IQR even if a significant disagreement was found between two groups.

Finally, the final value framework was created based on inclusion criteria: first, a simple majority threshold ($>50\%$) of all participants rating the indicator as *important* or *very important*; second, indicators must demonstrate stability across groups; and, third, indicators must have an IQR ≤ 1, indicating overall consensus among all participants. The final value framework was split into levels of "importance" based on the majority respondent result in R3. Appendix Table 4 shows an overview of the chosen methodologies and definitions.

Results

The exploration of this collaborative framework began with a proposal of 34 indicators based on the literature review, was adapted to 45 indicators based on participant contributions, and currently stands at 16 stable indicators with consensus created through value judgements in a Web-based Delphi exercise.

Web-Delphi Panel Participants

When organizing the Web-Delphi, 212 people were initially contacted to take part, of whom 129 accepted the invitation, including 25 patients, 32 HCPs, 29 supply-side actors, 19 decision makers, and 24 influencers. R1 had 101 active participants (78% participation rate), R2 had 91 active participants (70.5% participation rate), and by R3, there were 79 active participants, giving a 61% overall retention rate. A breakdown of participation and participant demographics is shown in Appendix Tables 7 and 8, respectively. Most participants were male, White, and between the ages of 30 and 60 years old.

Indicator Alteration from Round 1 Thematic Analysis

Of the initial 34 indicators proposed by the research team, 15 remained the same, 13 were altered and 6 were

Table 2 Interrater Agreement within Stakeholder and Country Groups across All Indicators

	Round 2			Round 3		
	Ky	95% CI		Benchmark Interval	Ky	95% CI		Benchmark Interval

	Ky	95% CI		Benchmark Interval	Ky	95% CI		Benchmark Interval
				Stakeholder groups				
Patients	0.72***	0.68	0.75	Substantial agreement	0.74***	0.70	0.78	Substantial agreement
Health care professionals	0.40***	0.32	0.48	Fair agreement	0.53***	0.45	0.61	Moderate agreement
Supply side	0.52***	0.42	0.63	Moderate agreement	0.59***	0.49	0.69	Moderate agreement
Decision makers	0.62***	0.56	0.67	Substantial agreement	0.66***	0.60	0.72	Substantial agreement
Influencers	0.58***	0.52	0.65	Moderate agreement	0.68***	0.61	0.75	Substantial agreement
				Country Groups				
Germany	0.43***	0.33	0.54	Fair agreement	0.48***	0.38	0.58	Fair agreement
United Kingdom	0.58***	0.50	0.65	Fair agreement	0.68***	0.61	0.75	Substantial agreement
United States	0.64***	0.59	0.69	Substantial agreement	0.61***	0.55	0.67	Substantial agreement

Source: The authors from the analysis of Delphi data. Interrater agreement measured by the Gwet's agreement coefficient with linear weights. Benchmark scale of the level of agreement as suggested by Landis and Koch[21]: Coefficient < 0.00, poor agreement; $0.00 <$ coefficient ≤ 0.20, slight agreement; $0.20 <$ coefficient ≤ 0.40, fair agreement; $0.40 <$ coefficient ≤ 0.60, moderate agreement; $0.60 <$ coefficient ≤ 0.80, substantial agreement; $0.80 <$ coefficient ≤ 1, almost perfect agreement [1].
***$P < 0.01$.

removed based on R1 qualitative feedback in the form of comments and respondent-added indicators (Appendix Table 9). Seventeen additional indicators were added to ensure all participant-added themes were incorporated in the draft framework, resulting in 45 indicators used in R2 and R3. Examples of indicators proposed by participants that were not included in the initial framework include "Multi-stakeholder design, development and implementation," "Sustainable data architecture," and "Integrates with and improves clinical processes."

Consensus Measurement

Agreement within groups. The IRA was calculated for each stakeholder and country group. Within the stakeholder groups, patients and decision makers maintained substantial agreement between R2 and R3, with HCPs being the least agreeable, moving from "fair agreement" (Ky = 0.40) in R2 to "moderate agreement" (Ky = 0.53) in R3 (Table 2). Within country groups, Germany had a limited movement toward consensus compared with the United Kingdom and United States, maintaining "fair agreement" throughout both R2 (Ky = 0.43) and R3 (Ky = 0.48) (Table 2). This indicates that British and American participants were more aligned in their value sentiments than Germans.

Stability between rounds. For an indicator to be unstable, a significant number of participants within a stakeholder group changed their answer between R2 and R3. Nineteen (out of 45) indicators were shown to lack stability in either country or stakeholder groups, with 16 indicators unstable in ≥ 1 stakeholder group and 12 in the ≥ 1 country group. Nine indicators overlapped between both stakeholder and country groups (see Table 3). Although there is a considerable amount of instability, in all unstable indicators, participants trended toward consensus and positivity; that is, participants were deciding whether indicators were "important" or "very important."

Stability among stakeholder groups. Concerning stakeholder groups, HCPs and influencers were the least stable, indicating a higher propensity to change their minds after seeing other participant responses. Their instability was predominantly in the "data rights and governance," "technical and security," and "clinical characteristics" domains. Examples include a significant number of HCPs changing their responses for "Sharing of identifiable data to outside commercial entities is not permissible" (Z = −1.995) and influencers for "Processes are in place to prevent unauthorized access to patient outcomes and data" (Z = −2.000) (Table 3). In both of these examples, many HCP respondents shifted their answers toward overall group consensus and higher importance ratings.

Stability among country groups. The United Kingdom was the most unstable country group, changing their mind significantly between rounds for 11 indicators, six of which were in the "data rights and governance" domain. The most unstable indicator for UK respondents was

Table 3 Significantly Unstable Indicators by Stakeholder Group and Country Group between R2 and R3

Value Domain	Indicator	Wilcoxon Test Z Value (>P Value)	Unstable Stakeholder Group
Data rights and governance	The storage and processing of data correspond to national legal requirements for data privacy	−1.997 (0.046)	Influencer
	Policies on data privacy, sharing, collection, and commercialization are clearly communicated to all users	−2.277 (0.026)	HCP
	Upon discontinuation of use, methods of de-registering and deleting data are clearly communicated to all users	−1.997 (0.046)	Influencer
	Systems are in place for health data (RWD) and its analytics (RWI) to contribute to real-world evidence (RWE) generation while adhering to privacy standards (including postmarketing approval)	−2.436 (0.015)	HCP
	Sharing of identifiable data to outside commercial entities is not permissible	−1.995 (0.046)	HCP
Technical and security	Complies with local data protections	−2.000 (0.046)	Influencer
	Security specifications are simply and transparently communicated to all users, including detailed information about updates	−1.996 (0.046)	HCP
	Processes are in place to prevent unauthorized access to patient outcomes and data	−2.000 (0.046)	Influencer
	Data interoperability	−2.441 (0.015)	HCP
	Capable of working and storing data offline and then syncing when internet is restored, where clinically appropriate	−2.228 (0.026)	Supply Side
	Where applicable, convenient, and sustainable device consumables	−1.995 (0.046)	HCP
Clinical characteristics	Evidence of clinical benefit based on patient-centered endpoints	−1.996 (0.046)	HCP
	Clinical risk management in place	−1.996 (0.046)	HCP
	Where applicable, improves quality of life for patients and carers	−1.997 (0.046)	Influencer
	Where appropriate, uses real-world data for proof of benefit	−1.995 (0.046)	HCP
User preferences	Where relevant, offers customizable integration with other solutions to facilitate management of multiple co-occurring conditions	−1.995 (0.046)	HCP

Value Domain	Indicator	Wilcoxon Test Z Value (>P Value)	Unstable Country Group
Data rights and governance	The storage and processing of data correspond to the regional legal requirements for data privacy	−1.988 (0.046)	UK
	Data are user owned	−1.987 (0.046)	UK
	Policies on data privacy, sharing, collectionm and commercialization are clearly communicated to all users	−2.000 (0.046)	UK
	Upon discontinuation of use, methods of de-registering and deleting data are clearly communicated to the user and are not difficult to achieve	−2.233 (0.026)	UK
	Systems are in place for health data (RWD) and their analytics (RWI) to contribute to real-world evidence (RWE) generation while adhering to privacy standards (including postmarketing approval)	−2.233 (0.026)	UK
	Sharing of identifiable data to outside commercial entities is not permissible	−1.997 (0.046)	UK

(continued)

Table 3 (continued)

Value Domain	Indicator	Wilcoxon Test Z Value (>P Value)	Unstable Country Group
Technical security	Data interoperability	−2.640 (0.008)	UK
	Capable of working and storing data offline and then syncing when Internet restored, where clinically appropriate	−1.997 (0.046)	Germany
	Sustainable data architecture	−1.977 (0.046)	UK
	Evidence of clinical benefit based on patient-centered endpoints	−1.988 (0.046)	UK
Clinical characteristics	Clinical risk management in place	−1.988 (0.046)	UK
Economic characteristics	Affordability to the patient	−1.988 (0.046)	UK

Source: The authors from the Delphi data analysis. Value aspects are displayed only if there is a significant difference ($P < 0.05$) in the median values for a set of groups between R2 and R3. Stakeholder group sample sizes are between 13 and 18 participants, and country group sample sizes are between 19 and 22.

Table 4 Indicators with Entire Group Dissensus in R3 Based on IQR

Value Domain	Indicators	Median	IQR
Health inequalities	Supports digital literacy	Important	2
	Helps reduce socioeconomic health inequalities	Important	2
Data rights and governance	Data are user owned	Very important	2
Technical and security	Uses multifactor authorization	Important	2
	Where applicable, ability for patient-users to input data	Neutral	2
Clinical characteristics	Supports and sustains lifestyle changes	Important	2
Economic characteristics	Premarketing approval, innovation incentives exist for supply-side actors, such as opportunities for managed entry and risk-sharing agreements	Important	2
	Affordability to the patient	Very important	2
User preferences	Connection to peer support where appropriate.	Important	2
	Provides an attractive/engaging experience for the end user	Important	2
	User can choose communication method as a result of personal preferences	Important	2
	Technology is offered in multiple languages	Important	2

Value aspects were included only if their IQR was >1 (i.e., there was no consensus among participants). The Likert scale has an interval of 1 between each answer.

"data interoperability" ($Z = -2.640$). Germany had one unstable response: "Capable of working and storing data offline and then syncing when internet is restored, where clinically appropriate" ($Z = -1.997$) (Table 3). In line with the overall response trends, these country group instabilities reflect individuals shifting their answers toward group consensus and high importance ratings.

Agreement across all respondents. IQR was used to measure consensus across all respondents. Sixteen were found to have stable consensus, ten being rated *very important* and six *important* according to most respondents (Appendix Table 10). Ten indicators had statistically significant stable dissensus (IQR = 2) across R2 and R3, three of which were in the user preferences domain ("Connection

to peer support where appropriate," "User is able to choose communication method as a result of personal preferences," and "Technology is offered in multiple languages") (Table 4). Two of the three health inequalities indicators also had a stable dissensus: "Supports digital literacy" and "Helps reduce socioeconomic health inequalities." The former has a negative rating of 25% and neutral rating of 18% in R3. Negative ratings (i.e., rating as *not at all important* or *little importance*) of indicators with dissensus ranged from 5% to 25%.

Subgroup Analysis

The Kruskal-Wallis H-test and Dunn's test were used to locate statistically significant group disagreements. Five

indicators were significantly disagreed upon between stakeholder groups; the most frequent disputes between supply-side actors and patients, with patients rating indicators significantly more positively (of higher importance) than the supply-side (Table 5). HCPs significantly disagreed with both influencers and patients indicator "Helps reduce socioeconomic health inequalities" with 6% of HCPs rating the indicator as *very important* compared with 44% of influencers and 60% of patients.

Stakeholders from Germany and the United States disagreed significantly on four indicators in the health inequalities, technical and security, and clinical characteristics domains. The indicator "Supports digital literacy" was opposed, with 0% of Germans rating it as *very important* compared with 37% of Americans (Table 5). Table 6 includes all indicators and displays stability across R2 and R3 and consensus across stakeholder and country groups. Overall rated importance reflects the percentage of respondents that rated the indicator as either *important* or *very important* in R3. Indicators range from 48% to 100% rated overall importance.

Final Value Framework

The 16 indicators with stable consensus that comprise the value framework (Table 6) reflect ideas around a system's willingness to pay, the potential for health system integration, the user experience, and technical reliability. A system's willingness to pay includes value indicators such as affordability, cost-effectiveness, sustainability, and value-based care. The potential for system integration focuses on clinical processes and patient activation, while the user experience focuses on indicators such as ease of adoption, user support, multistakeholder design, and user engagement. Technical reliability has to do with the performance of the technology, including regulatory compliance, procedures for software updates, and well-established data provenance.

Discussion

Value assessment frameworks are essential building blocks for HTA methods that traditionally rely on assessing the efficacy and cost of new technologies, while assessment of societal impact comes second. This approach is changing, though, as Multiple Criteria Decision Making (MCDA) methods gain increased popularity and health systems shift toward value-based care models. The penetration of digital technology in health care is also a significant factor driving this shift, because data collection creates broader impacts and raises assessment issues beyond those previously raised by

pharmaceuticals and medical devices. In this Delphi study, we solicited value judgments across various domains and indicators, including those specific to digital technology and data collection. In line with sentiments throughout the digital health space, our results indicate that stakeholders are not entirely confident of where value lies in provider-facing DHTs and how they should be assessed.

Ultimately, this exercise aims to help guide decision makers in their duty to translate from stakeholder value sentiments into reimbursement decisions. The resultant value framework can be operationalized and adapted to suit different contexts within the provider-facing DHT space. All stable indicators with consensus are represented in the framework, but some have higher importance scores than others do; thus, indicators with high importance scores may correspond to stronger recommendations for reimbursement considerations. In addition, not all indicators will be linkable to reimbursement. Some indicators in this framework are more relevant for legal, ethical, and regulatory considerations, such as information governance standards.

As illustrated in the "Results" section, there were high rates of instability across stakeholders and value domains. Instability is a notable finding on its own, as it reflects changing value judgments and uncertainty from stakeholders. For several indicators, at least one group changed their mind significantly between the second and third rounds of the Delphi exercise. In fact, only one indicator in the "data rights and governance" domain was stable enough to draw conclusions. Interestingly, this one indicator, "Adheres to strong information governance standards, including processes for data protection violations," had the highest rating of all indicators, with 85% of all participants rating it *very important* in R3. This instability demonstrates the novelty of provider-facing DHTs and digital impact on the health system at such a large scale. Assessing these types of DHTs for chronic disease management is truly a new challenge for all health system stakeholders, as there is wide variety in DHT design and purpose. With much left unknown, stakeholder participants seemed to defer judgments to others and changed their answers to align with the majority. However, it is important to note that, of the unstable indicators, the predominant movement was between the ratings of *important* and *very important*, indicating that participants acknowledge the relevance and importance of the indicators. All unstable indicators except one showed shifting judgments to group consensus and higher importance ratings. Only one indicator, "Capable of working and storing data offline and then syncing when Internet is restored, where clinically appropriate," did not

Table 5 Indicators in R3 with Significantly Different Group Responses

Value Domain	Indicator	Kruskal-Wallis H Test[a]	Stakeholder Groups: Dunn's Test for Pairwise Comparisons[b] and Percentage of Group Positive Judgments[c]
Health inequalities	Helps reduce socioeconomic health inequalities	12.761**	HCP (56%) – Influencer (94%)** HCP (56%) – Patient (87%)***
Technical and security	Where applicable, ability for patient-users to input data	9.951**	Supply (35%) – Patient (24%)**
Clinical characteristics	Improves patient adherence	11.329**	Supply (53%) –Patient (100%)***
	Supports and sustains lifestyle changes	10.597**	Supply (41%) – Patient (87%)***
	Where applicable, improves quality of life for patients and carers	10.477**	HCP (69%) – Patient (100%)***

Value domain	Indicator	Kruskal-Wallis H test[a]	Country groups – Dunn's test for pairwise comparisons[b] and percentage of group positive judgments
Health inequalities	Supports digital literacy	10.012***	Germany (33%) – United States (79%)***
Technical and security	Where applicable, convenient and sustainable consumables	6.148**	Germany (52%) – United States (79%)**
Clinical characteristics	Improves patient adherence to treatment	7.375**	Germany (71%) – United States (95%)***
	Supports and sustains lifestyle changes	11.774***	Germany (43%) – United States (84%) ***

For instability, value aspects are displayed only if there is a significant difference in the median values for a country between round 2 and 3. Sample size ranges from 19 to 22.

[a]Chi-squared adjusted for ties with 4 degrees of freedom. For simplicity of the table, only the value aspects for which a statistically significant result was found are showed.

[b]Dunn's test (pairwise multiple comparisons using the Bonferroni adjustment) **$P < 0.05$; ***$P < 0.01$. Percentages display the percentage of each country group that rated each indicator positively, that is, *important* or *very important*.

[c]Positive judgments include participant ratings of *important* or *very important*.

change ratings to the R2 consensus of *very important* but still showed increasing importance ratings, with most rating *important*. Overall, the change trended toward group consensus whenever participants changed their answers. Further rounds of Delphi are needed to establish the stability of these indicators to determine a conclusive rating of *important* or *very important*. A quarter of all indicators demonstrated dissensus among all respondents (Table 4), four of which also had significant group disagreements (Table 5). Dissensus resulted from high neutrality ratings and considerable distribution of responses among the importance scale (distribution of responses can be seen in Appendix Figure 2). Although there was a minor movement toward consensus and positivity, there was limited movement of responses between the rounds, resulting in the inability to make conclusions. This minimal rating movement indicates that participants were not budging in their decisions. Thus, without gaining an insight into why these disagreements were happening, further rounds of Delphi may be unable to help participants reach

consensus. To move forward and understand the reasoning behind this dissensus, conversations between different respondents and stakeholder groups could be facilitated to understand why decisions were made and what precisely the disagreements were. For example, the indicator discussing premarketing approval had little to no movement between the rounds, maintaining 32% neutral ratings. For highly complex topics, such as premarketing approval, it would be particularly impactful to discuss the topic in detail among different stakeholders to understand where disagreements are happening and ensure a solid base-level understanding of the topic to facilitate discussions surrounding their importance related to provider-facing DHTs.

Wide-scale digital innovation is highly complex, and all stakeholders involved must take the necessary time to learn about and understand these new technologies to be able to interpret what kind of impact they may create and how they may offer value. For this reason, knowledge-sharing and co-creation initiatives such as

Table 6 Final Value Framework for Provider-Facing DHTs Showcasing Ranked Accepted Indicators according to the Rating of Most Participants

Value Domain	Ranked Stable Indicators with *Very Important* Rated by Majority	Percentage of Participants Rating *Very Important*
Data rights and governance	Adheres to strong information governance standards, including processes for data protection violations	85%
Economic characteristics	Long-term cost-effectiveness to the system	81%
Technical and Security	Data has a high degree of integrity and credible provenance	78%
Economic characteristics	Affordability to the system	72%
Clinical characteristics	Integrates with and improves clinical processes	71%
User preferences	Relevance of the solution to the targeted user group	62%
Technical and security	Systems are in place for continued product development and security updating after product release	57%
Clinical characteristics	Improves communication and information sharing	47%
Economic characteristics	Adheres to value-based care methodology	47%
Health Inequalities	Does not exacerbate existing health inequalities	44%

Value Domain	Ranked Stable Indicators with *Important* Rated by Majority	Percentage of Participants Rating *Important*
User preferences	Technical and user support	51%
Economic characteristics	Sustainable system improvements through resource optimisation	46%
User preferences	Ease of adoption and use with minimal training	47%
Clinical characteristics	Improves personal health engagement and patient activation, including improved decision-making abilities	43%
User preferences	Multi-stakeholder design, development, and implementation	46%
User preferences	Provides an attractive/engaging experience for the end-user	44%

this one are crucial to the success of this space. Perhaps there is not a comprehensive understanding within and across stakeholder groups of exactly the magnitude of impact these digital innovations can impose on health systems. The health sector has been largely shielded from the effects of Big Data integration due to regulatory protections. However, we have reached a critical juncture in digital health development that narrowing these knowledge gaps has become an urgent initiative. Two of the three indicators in the health inequalities domain had dissensus based on IQR and significant stakeholder group disagreements: "Reduces socioeconomic health inequalities," with disagreements between HCPs and influences, and HCPs and patient-users, and "Improves digital literacy," with disagreements between Germans and Americans (see Table 5). Significant disagreements between stakeholder groups highlights the importance of a multistakeholder collaboration approach that uses the Delphi method to build consensus. Particularly regarding social value indicators around health disparities, including participants from a diverse set of health backgrounds ensures that a wide variety of viewpoints are represented. DHTs can both mitigate and perpetuate health inequities and the digital divide. Still, to improve

the current situation, they must use user-friendly interfaces and AI trained on unbiased data. AIs have a unique opportunity to reduce health inequities, such as the ability to identify health disparities in marginalized communities and the development of targeted interventions. On the other hand, AI is susceptible to algorithmic bias and can certainly lead to an increase in inequity, particularly racial bias, if there is poor representation in its training data set, human errors in design, inaccurate research questions, and/or data set shift, which is why it is so important to have methods in place to evaluate AI-based DHTs.[24] Bad outcomes can also occur with inadequate user involvement throughout development, potentially resulting in a user experience that is not aligned with the needs of the primary user group. It would be helpful to confirm whether all participants were aware of these dynamics. Participants were screened for eligibility based on their knowledge of and/or experience with DHTs, but individual expertise varied across participants.

Despite the novelty and complexity of assessing of DHTs, specifically provider-facing DHTs that can affect more than one patient and involve multiple types of technologies, we found there to be consensus and stability across several criteria, ten of which the majority of

respondents rated as *very important* and six as *important*. The three highest-rated rated indicators were "Adheres to strong information governance standards, including processes for data protection violations," "Long-term cost-effectiveness to the system," and "Data has a high degree of integrity and credible provenance," respectively, rated 85%, 81%, and 78% *very important* for the assessment of provider-facing DHTs. The concepts present within these indicators are important in enabling health systems to move toward efficient, proactive, and collaborative care, which provider-facing technologies have strong capabilities to enable. Strong information governance standards and data integrity are essential for merging data sets, maintaining data privacy and security across settings and integrating data into electronic health records. These indicators also highlight the diligence respondents place on data, specifically on ensuring its accuracy and reliability, and how technologies must be responsible for data privacy and security.

Implications

We believe there is a potential for significant implications of this research. Our findings indicate a need for continued investigations in this area to implement HTA policies that accurately evaluate digital health innovations. These issues also raise the question of whether current regulatory frameworks must be updated to ensure the privacy and security of health information, particularly in an age of personalized medicine. While some indicators can be part of HTA appraisals, others may require institutional intervention. Among other issues, widespread data collection and RWE to meet evidence standards illustrate that DHTs not only bring forward new domains of value to be assessed but also point out gaps in regulatory needs. By understanding what different stakeholders value, health care decision makers can introduce policies promoting the creation of solutions that can meet the needs of multiple stakeholders.

Limitations

Our study is not without limitations. First, additional Delphi rounds will need to be conducted to establish stability for all indicators to draw complete conclusions. A fourth Delphi round might have reduced the number of indicators with significant instability. Second, Delphi R1 did not have an option for participants to alter value domains, only indicators, which may limit creativity and introduce bias during the open-ended portion of the panel. However, the starting framework resulted from an extensive literature review coupled with expert input from

stakeholders. Third, results from this exercise are not generalizable to all health systems as they pertain to three specific countries; nonetheless, they provide a helpful starting point for a similar exercise across several contexts. Fourth, during recruitment, some, not all, participants were paid by professional recruiters to participate to ensure an equal number, and therefore a comparable sample, of participants from our three study countries and five stakeholder groups. The research team decided that the importance of equal representation outweighed potential biases that professional recruiters may have introduced. There is an ongoing debate in the research community about the type of bias introduced by paying or not paying participants; please see Appendix section 2.2 for further information. Finally, it is not possible to conduct power analysis for nonparametric tests in STATA. Therefore, subgroup analysis regarding disagreements between stakeholder and country groups were not powered, and effect sizes could not be calculated.

Conclusion

We have created a value assessment framework for the assessment of provider-facing DHTs comprising a total of 45 indicators, 16 of which were stable, resulting in consensus among participants. There was some instability in participant responses, particularly regarding indicators within the "data rights and governance" domain. To a certain extent, this is to be expected due to the novelty of the integration of data and analytics in heath technologies and their assessment. Furthermore, the impact of these technologies beyond single patients adds an additional layer of complexity to the technology assessment process. Further work is needed to ensure that, across all types of stakeholders, there is a clear understanding of the potential impacts of system-facing DHTs. Such far-reaching and complex impacts highlight the importance of continued multistakeholder involvement in the development of value assessment frameworks.

ORCID iD

Caitlin Main https://orcid.org/0000-0001-5157-5283

Supplemental Material

Supplementary material for this article is available online at https://doi.org/10.1177/0272989X231206803.

References

1. Oortwijn W, Sampietro-Colom L, Habens F. Developments in value frameworks to inform the allocation of healthcare resources. *Int J Technol Assess Health Care.* 2017;33:323–9.
2. Faulkner E, Holtorf A-P, Walton S, et al. Being precise about precision medicine: what should value frameworks incorporate to address precision medicine? A report of the personalized precision medicine special interest group. *Value Health.* 2020;23:529–39.
3. Kidholm K, Ekeland AG, Jensen LK, et al. A model for assessment of telemedicine applications: MAST. *Int J Technol Assess Health Care.* 2012;28:44–51.
4. Pichon-Riviere A, Soto N, Augustovski F, Garcia-Marti S, Sampietro-Colom L. Involvement of relevant stakeholders in health technology assessment development. *Int J Technol Assess Health Care.* 2018;34:248–53.
5. Kamal RN, Lindsay SE, Eppler SL. Patients should define value in health care: a conceptual framework. *J Hand Surg Am.* 2018;43:1030–4.
6. National Institute for Health and Care Excellence. Evidence standards framework for digital health technologies: user guide. 2021;19. Available from: https://www.nice.org.uk/corporate/ecd7/chapter/introduction
7. Guo C, Ashrafian H, Ghafur S, et al. Challenges for the evaluation of digital health solutions—a call for innovative evidence generation approaches. *NPJ Digit Med.* 2020;3:1–14.
8. MedTech Europe. Recognising the value of digital health apps: an assessment of five European healthcare systems. Available from: https://www.medtecheurope.org/wp-content/uploads/2021/11/2111_v4.8_mte_dht_reimbursement16.11.2021-2.pdf
9. Kruse CS, Goswamy R, Raval Y, et al. Challenges and opportunities of big data in health care: a systematic review. *JMIR Med Inform.* 2016;4:e38.
10. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches. *Am J Prev Med.* 2016;51:843–51.
11. Manteghinejad A, Javanmard SH. Challenges and opportunities of digital health in a post-COVID19 world. *J Res Med Sci.* 2021;26:11.
12. RWE Navigator. Sources of real-world data—RWE Navigator. Available from: https://rwe-navigator.eu/use-real-world-evidence/sources-of-real-world-data/. [Accessed 25 July, 2023].
13. Oh W, Nadkarni GN. Federated learning in health care using structured medical data. *Adv Kidney Dis Health.* 2023;30:4–16.
14. Sharma B, Halder R, Singh J.Blockchain-based Interoperable healthcare using zero-knowledge proofs and proxy re-encryption. In: *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS).* Bengaluru (India): IEEE; 2020, pp. 1–6.
15. 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 Health.* 2023;26:1474-84.
16. Sauermann S, Herzberg J, Burkert S, Habetha S. DiGA – A chance for the German healthcare system. *J Eur CME.* 2022;11:2014047.
17. von der Gracht HA. Consensus measurement in Delphi studies: review and implications for future quality assurance. *Technol Forecast Soc Change.* 2012;79:1525–36.
18. Holey EA, Feeley JL, Dixon J, et al. An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC Med Res Methodol.* 2007;7:52.
19. Welphi. Home. Available from: https://www.welphi.com/en/Home.html. [Accessed 30 November, 2022].
20. Lumivero. NVivo. 2020. Available from: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home
21. Gould W. StataSE. 2019. Available from: https://www.stata.com
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Int Biometric Soc.* 1977;33:159–74.
23. Hirschhorn F. Reflections on the application of the Delphi method: lessons from a case in public transport research. *Int J of Soc Research Methodol.* 2019;22:309–22.
24. National Institutes of Health. Ethically leveraging digital technology for health. *An Examination of Emerging Bioethical Issues in Biomedical Research: Proceedings of a Workshop.* 2020. Available from: https://www.ncbi.nlm.nih.gov/books/NBK560366/