



OPEN Development of an evidence-based evaluation framework for digital health software products

Jennifer C. Goldsack¹, Colby Holliday¹, Yashoda Sharma¹, Doug Mirsky¹, Cole Zanetti^{2,3} & Benjamin Vandendriessche^{1,4}✉

Current frameworks and standards for evaluation of digital health software products (DHSPs) are fragmented and may not cover the spectrum of what stakeholders consider most important. We conducted a mixed-methods study with the goal of producing an evidence-based and open-access evaluation framework (EF). We conducted a needs assessment among 173 subject matter experts (SMEs) to assess the need for this framework and identify its components using interviews, focus-group discussions, and a survey. Next, we conducted a narrative review of 1053 papers, and a landscape analysis of 160 relevant frameworks, guidances, and standards. This data was used to develop the components of the EF: high-level themes (domains), with evaluation criteria and associated benchmarks for evaluating the quality and trustworthiness of DHSPs. An additional 58 SME interviews were conducted to validate the EF. The top three domains for assessing overall quality were evidence, usability, and privacy and security. Equity was a common theme across all domains. The EF spans multiple domains of trust and value, harmonizes best practices, stands on the shoulders of well-respected and established work, and eases the effective adoption of high-quality, trustworthy DHSPs.

Keywords Digital health technologies, Evaluation framework, Health software, Best practices

Digital health software products (DHSPs) are increasingly used by patients, caregivers, and healthcare professionals in the delivery of care to manage, maintain, or improve health¹. DHSPs are software applications built for a general-purpose computing platform, standalone products, extensions to another standalone product, or companions to hardware sensors.

Despite the rapid uptake of DHSPs and the recognition of their potential value in healthcare, many question how the quality of these technologies is assessed and are calling for more evidence-based evaluation frameworks^{2,3}. Borges et al. recently identified lack of infrastructure and technical support, impact on clinician workload, inadequate training, and perception of usefulness as barriers to healthcare provider's willingness to adopt DHSPs⁴. Clinicians and patients struggle with identifying the clinical utility of DHSPs^{5,6}. Some clinicians have shared that they are hesitant to recommend DHSPs to patients due to lack of knowledge on identifying which ones are effective or can be trusted⁷. This disconnect may stem from gaps with clinicians' limited knowledge of medical product approvals in general; in a recent US survey, only 17% of physicians indicated some level of understanding of the FDA's device approval process⁸.

The current process of establishing that DHSPs are built according to best practices is haphazard and fragmented⁹. Numerous standards, guidances, and audits exist¹⁰, covering domains such as evidence (e.g., FDA de novo, 510(k)), privacy and security (e.g., HITRUST, SOC2) and usability (e.g., WCAG, HFE/UE report), but none covers the full spectrum of what users may consider important. Moreover, where regulatory requirements do not exist, industry evaluations, which can be very expensive and time consuming for developers, are left up to the developer to consider. Indeed, many consumer health and wellness DHSPs are not regulated because they do not meet the definition of a medical device, leaving the decision to audit a product up to the developer, and consumers the task to ascertain quality and trust of a product, often without expertise to do so. As a result, many potential adopters of DHSPs—especially large healthcare systems, insurers, and payers—develop their own bespoke evaluation flows, leading to further fragmentation¹¹.

To that end we undertook to develop a single framework to establish that products are built according to best practices and achieve a common baseline of acceptability, to speed the adoption of valuable digital health

¹The Digital Medicine Society, Inc. (DiMe), 90 Canal St 4th Fl, Boston, MA 02114, USA. ²U.S. Department of Veterans Affairs, Charleston, SC, USA. ³College of Osteopathic Medicine, Rocky Vista University, Parker, CO, USA. ⁴Department of Electrical, Computer, and Systems Engineering, Case Western Reserve University, Cleveland, OH, USA. ✉email: benjamin.vandendriessche@dimesociety.org

technologies (DHTs) and build trust among buyers and end-users across the healthcare landscape. To do so, we conducted a 2-phase iterative study.

In phase 1, we interviewed and surveyed subject matter experts (SMEs), developed a needs assessment, and identified the high-level components of an evaluation framework (EF). In phase 2, we cataloged the current state of science and regulatory guidances to create an evidence-based EF that comprises multiple domains of interest, each of which is composed of a set of evaluation criteria and associated benchmarks. The ultimate goal was to develop an EF that DHSP adopters can use to ensure that their products reflect best practices and meet a quality bar that instills trust.

Data and methods

This mixed-methods study collected evidence to develop an EF for a common baseline of acceptability for DHSPs. Evidence was collected in phases that informed each other. All methods were carried out in accordance with relevant guidelines and regulations, and the experimental protocols were approved by the Advarra institutional review board (Advarra Pro00073478). Informed consent was obtained from all the participants, prior to interviews, focus groups, and surveys to use anonymized answers for the purpose of this work.

Needs assessment

The needs assessment elucidated the need for a framework to evaluate the quality of DHSPs (SFig 1). We conducted interviews and focus groups with SMEs representing stakeholders from across the healthcare ecosystem. From August to October 2023, we recruited 164 potential participants from DiMe's network of digital health experts, including from regulatory agencies, healthcare providers, DHSP developers, patient advocacy, medical societies, life science companies, payers and investor organizations, of which 79 (48%) agreed to participate. We specifically focused on English-speaking individuals in mid- to senior-level positions; 40% were women and the geographical location of the organizations' headquarters was: 76% US, 7% Europe, 4% elsewhere, and 13% with a global footprint.

All interviews ($n=50$) and focus groups ($n=29$ participants) were held via Zoom except one focus group which was held in person. Interviews lasted 45 min and included 1 participant, 1 study facilitator, and 1 note-taker. Interviews were semi-structured and conducted in two cohorts. In the first cohort, 35 interviews were conducted to identify the high-level topics (i.e., domains) a comprehensive framework should include. These domains were validated by the second cohort of interviews ($n=15$). In addition, both cohorts were asked to discuss key trends in how DHSPs are evaluated across the industry, and to identify participants' greatest needs and priorities for an evaluation program for their specific stakeholder group.

Focus groups consisted of 8–13 participants and lasted 1 h; 1 group consisted only of DHSP adopters ($n=9$, healthcare providers, payers, and other stakeholders that aggregate product information and recommendations), 1 consisted only of DHSP developers ($n=8$), and the remaining group, which met in person, consisted of adopters, developers, and industry associations and investors ($n=13$). An online workspace¹² was used during the focus groups; this allowed participants to share additional comments during the discussion. Participants in all 3 focus groups were asked the following questions: (1) What EFs, certifications, or standards are you aware of or currently using?, and (2) What value would an EF for DHSPs bring to healthcare? In the adopters-only focus group, we also asked how developers could make the process of choosing a DHSP more efficient. In the developers-only group, we asked how adopters and regulators could make their work easier.

Survey

From the information gleaned from the interviews and focus groups, we developed a survey with 76 questions based around the domains identified in the interviews and focus groups (SFig 1). From October to November 2023, we shared the survey broadly to the digital medicine community through DiMe's partner email list and Slack community.

We evaluated content validity¹³ with SMEs to ensure that the survey questions represented concepts that are essential for assessing the quality of DHSPs, and we asked experts ($n=15$) representing developers, adopters, investors, and regulators to review the survey questions' validity. Based on recommendations¹⁴, a panel of ≥ 5 participants is sufficient to assess the quantified content validity ratio (CVR). The panelists were asked to specify whether a question was "not necessary," "useful but not necessary," or "essential". The content validation index (CVI) was then compared with the Lawshe critical value¹⁵; a critical value of 0.99 was needed for validation. A total of 17 questions were removed as a result of this process, leaving 59.

Evaluation framework development

We conducted a literature review, guided by findings from the needs assessment, to benchmark how quality has been assessed for DHSPs (SFig 2). Its aim was to identify quality assessment criteria that would meet the needs of both developers seeking to differentiate their products as trustworthy and adopters identifying which products are worthy of further consideration.

We set the inclusion and exclusion criteria (SFig 3) to identify publications that included recommendations for assessing DHSPs within the parameters set by the needs assessment. We screened 4504 English-language titles and abstracts published between January 2020 to December 2023 in PubMed, Google Scholar, and Publish or Perish¹⁶ against the inclusion and exclusion criteria, and extracted data from 1551 publications that met these criteria. Titles and abstracts were independently screened by 5 researchers. A subset of 10% was randomly selected for auditing by 1 researcher who was blind to the initial screening judgment. Cases of disagreement were discussed and resolved by consensus of all 5 researchers. Using findings from the needs assessment, we applied a combination of deductive and inductive thematic analysis to the extracted data¹⁷. Five researchers coded data for domains, intended users, and product types; a DHSP discussed in the publication was assigned to one of 3

intended user groups (patients or consumers only, patients and clinician team, or clinical or administrative staff only). We assigned each publication to 1 or more domains and 1 or more product types, based on the information provided about the DHSP. Additionally, the researchers applied inductive analysis to identify descriptive themes around quality assessments within the three domains¹⁸.

Secondly, we conducted a landscape analysis, following the WHO guide¹⁹, to comprehensively identify frameworks, guidances, and standards that apply to DHSPs. The goal of this analysis was to extract recommendations and best practices that could be used to design a comprehensive EF for DHSPs. The sources were identified through web searches, guided by knowledge from DiMe and regulatory experts on which sources are considered important by the field for evaluating DHSPs. The identified sources were labeled as a certification, framework, guideline, industry standard, regulatory guidance, or tool. Next, the identified sources were assigned to at least one of the domains. During interviews (see below), additional sources were identified iteratively until saturation was reached.

Findings from the literature review and landscape analysis served as the basis for the EF. We developed domain-specific criteria groups, and criteria, and benchmarks that could be used to evaluate the quality of DHSPs. From March to April 2024, we interviewed 49 additional SMEs, including patient representatives, to validate the criteria and benchmarks. Again, we engaged stakeholders from across the healthcare industry. Interviews were held via Zoom. We shared the criteria groups and criteria with participants before interviews; during interviews, we reviewed each benchmark in the context of the domains and criteria, and collected feedback. We asked participants to evaluate the relevance of each criterion and benchmark.

Finally, in May 2024, we conducted user testing interviews with DHSP developers to validate the feasibility of attesting a product against the identified criteria and benchmarks. Saturation was reached after 9 interviews.

Results

Needs assessment

The first cohort of SME interviews identified usability, equity and inclusion, clinical and technical evidence, market and end-user evidence, and privacy and security as the major high-level topics, i.e., domains. Additionally, these interviews identified 4 stakeholder groups that would stand to benefit from a comprehensive EF: DHSP adopters (including healthcare systems, payers, and patients), DHSP developers, regulators, and industry associations.

For each interview in the second cohort, we sorted feedback by domain and employed an inductive approach for thematic analysis. The feedback received quickly reached saturation, as the participants shared similar needs and challenges that could be addressed with an EF. We moved to analysis after 15 interviews.

The domains were validated according to how many responses corresponded to a given domain or the number of comments that included a detail relevant to that domain. For adopters, usability received the most comments, followed by evidence, equity & inclusion, and privacy & security. Respondents did not separate out market and end-user evidence from clinical and technical evidence, but instead spoke of evidence as a single domain. Developer responses ranked the domains in a slightly different order: evidence, usability, privacy & security, and equity & inclusion. This analysis confirmed that the domains identified in the first cohort covered the areas that adopters and developers focus on when vetting or designing DHSP.

We used an inductive approach to review responses within each domain to identify themes and key trends for evaluation of DHSPs, spanning responses from adopters and developers (STable 1). This analysis also identified workflow integration, outcomes, and the overall business model as relevant context for assessing the quality of DHSPs; SMEs preferred that the domains serve as the central organizing unit for the EF.

The top three domains that both adopters and developers wanted to prioritize were evidence, usability, and privacy & security, with equity & inclusion as a 4th theme that applies to all domains. The themes for evidence varied slightly; adopters prioritized evidence vetted by clinicians and supporting workflow integrations, whereas developers prioritized evidence to support clinical claims and ROI. For usability, both stakeholder groups prioritized demonstrating knowledge of user needs for and value of using the DHSP. For privacy & security, adopters prioritized clearly defined measures without specifying a particular method, whereas developers stated that the reference standards of HITRUST²⁰, SOC 2 Type II²¹, and HIPAA²² should be in place (STable 1).

We developed content for focus group discussions around the finding that three domains are foundational. We began these discussions with prompts to learn about the need for an EF for developers and adopters. We then asked participants to assess each domain in the context areas: outcomes, equity & inclusion, workflow, and business model. We also asked participants to identify any gaps they believe exist in the ability of adopters to efficiently evaluate DHSPs. Participants were aware of or are currently using numerous frameworks, certifications, and standards for the privacy & security domain (such as HITRUST, HIPAA, SOC 2 Type II, SMART on FHIR, HITECH, SaMD, NCQQ, VA/DoD, FedRAMP, ONC, KLAS reports, GDPR)^{20–30}. Fig. 1

We then presented the adopter and developer focus-group participants with a grid depicting the domains and context areas. When presented with the question, “What value would an EF for DHSPs bring to healthcare?”, participants’ answers were grouped thematically (Fig. 2; STable 4).

Across the needs assessment activities, participants were asked for each domain if they believe “good exists and there is very little need for improvement” or “the current state is insufficient and there are many opportunities for improvement” (Fig. 1 and STable 2).

Survey development and deployment

We synthesized findings from the interviews and focus groups to design a survey aimed at quantifying the impact of the gaps identified in the focus groups. Specific questions were developed around the three domains.

The survey was sent to the DiMe community and completed by 93 participants: 45 adopters, 32 developers, and 16 regulators, investors, and industry association representatives. The top roles represented were executive leadership, research, data science, or analytics profiles (STable 5). All 3 groups ranked “clinical outcomes clearly

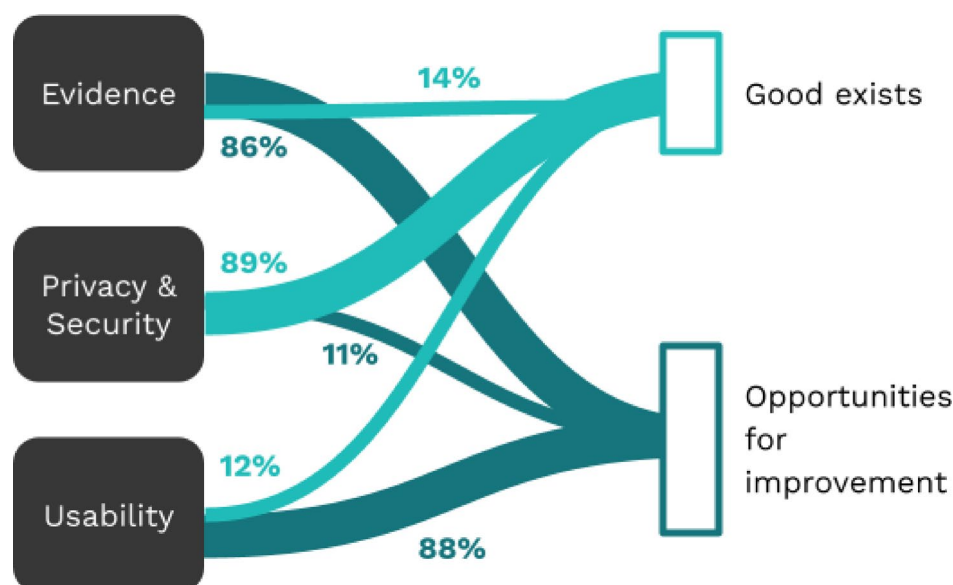


Fig. 1. Representation of SME answers to the question “In what domain does good exist?” and “In which domains is there room for improvement?” to evaluate DHSPs for quality and trust. Respondents: evidence, $n = 36$; privacy & security, $n = 18$, usability, $n = 26$.

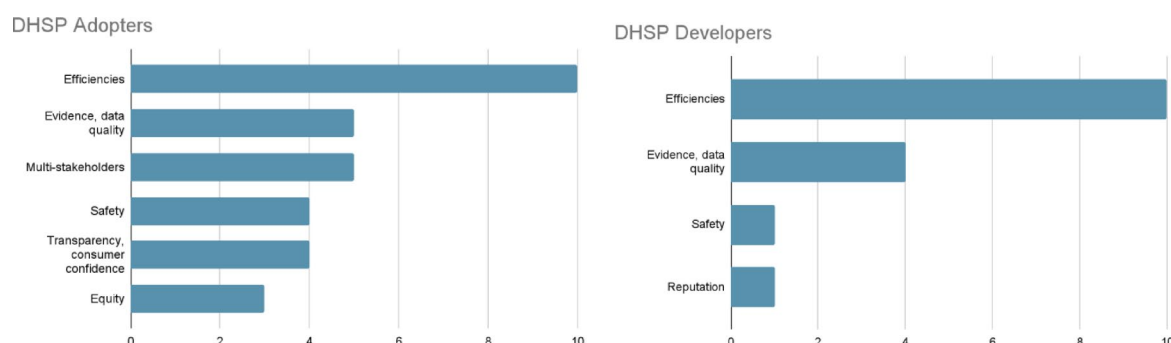


Fig. 2. Thematically grouped responses to the question “What value would an evaluation framework for DHSPs bring to healthcare?”. (left) Total times a theme was identified (after thematic analysis) in the DHSP adopter focus group; (right) The same data for the DHSP developer focus group. In order of appearance: improved efficiencies, improved data quality and evidence, generate multistakeholder agreements, improved safety, improved transparency leading to improved consumer confidence, improved equity, improved reputation (of product or company).

defined” as the top criterion when evaluating a new DHSP (Table 1). All 3 groups also ranked “easier to tell which products are fit for my purpose” as the most valuable aspect of such a framework. The evidence domain was ranked as most important when evaluating DHSPs, and “outcomes” was ranked far above equity & inclusion or workflow integration when asked about context.

Evaluation framework development

Several outcomes from the needs assessment were used for the next research phase. This included condensing to 3 domains—evidence, privacy/security, and usability—for 4 stakeholder groups: adopters, developers, regulators, and industry associations. The theme of equity was ubiquitous and woven throughout the domains. The thematic analysis and survey results provided content for criteria to evaluate each domain. We also organized DHSPs into 3 user types based on the intended user group: patients or consumers only, patients and clinical teams, and clinical teams or administrators only.

For the literature review, we screened the titles and abstracts of 4504 unique publications. We reviewed the full text of 1551 (34%), of which 1053 (68%) provided recommendations or best practices to evaluate DHSP quality (SFig 2). We extracted information from these publications that could inform DHSP quality within the defined domains and identified the intended user group for the subject DHSP.

Through inductive thematic analysis, we identified three recommendations for improving quality for the evidence domain: engage a variety of stakeholders when developing or validating a DHSP, conduct a research

Responses	Adopters (n = 45)	Developers (n = 32)	Regulators, Investors, Industry associations (n = 16)
<i>Question: Which of these criteria is most important to you when evaluating a new DHSP?</i>			
Clinical outcomes clearly defined	27%	24%	26%
Appropriate privacy/security measures	23%	23%	10%
Ease of use	18%	23%	16%
Product integrates into workflows	16%	14%	13%
Inclusively designed, where appropriate	5%	13%	10%
Overall return on investment	8%	3%	19%
Other	3%	0%	6%
<i>Question: What value would an evaluation framework for DHSPs bring to healthcare?</i>			
Easier to tell which products are fit for my purpose	24%	20%	25%
Clear picture of costs, reimbursement options	18%	10%	14%
Faster time to adoption of new products	17%	19%	18%
Better products in the market	16%	17%	14%
Increased transparency for decisions made during product development process	15%	20%	18%
More innovative products in the market	10%	14%	11%
No value	0%	0%	0%
<i>Question: Which domain is most important when evaluating a new DHSP?</i>			
Evidence	64%	43%	64%
Usability	25%	33%	18%
Privacy & security	11%	24%	18%
<i>Question: Which context area is most important when evaluating a new DHSP?</i>			
Outcomes	67%	62%	64%
Equity and inclusion	19%	19%	36%
Workflow integration	14%	19%	0%

Table 1. Survey responses.

study for evidence generation, and build consensus around evidence guidelines (STable 3). Within the domain of privacy & security, there was a call for more comprehensive guidelines specific to DHSPs, and more resources and information on data privacy for end-users. The themes we identified for the usability domain were also focused on providing more information and conducting user testing.

The landscape analysis revealed 160 professional sources for alignment with the three domains, product types, and recommendations or best practices that could inform an EF (STable 6): 47 regulatory guidances, 32 frameworks, 32 guidelines, 34 industry standards, and 15 tools. From these, 92 provided information relevant to the evidence domain, 81 related to privacy and security, and 62 related to usability.

Data from the needs assessment, survey, literature review, and landscape analysis were integrated into an EF organized around each domain. For each domain, we identified and defined criteria groups as the first level of organization (Table 2). We further divided each criteria group into criteria and associated benchmarks.

We interviewed SMEs to validate and refine this initial version of the EF. We asked them to approach their review of the framework through the lens of the stakeholder group most closely tied to their role. After interviewing 49 SMEs (16 adopters, 21 developers, and 12 industry association representatives) and observing saturation with the feedback, we moved to analysis and synthesis to integrate the feedback in the framework.

SMEs from all stakeholder groups expressed that the framework was comprehensive and contained important details. They recommended approaches to attesting to the benchmarks that ranged from federal regulations such as FDA 510(k) clearance to a document summarizing the work conducted or processes in place. They thought that the evidence domain could benefit most from an EF, as very few requirements exist for what good evidence should look like, especially for products that are not subject to FDA oversight. The primary feedback for privacy & security was that many privacy and security regulations exist for DHSPs. For usability, the primary feedback was that the usability criteria are important and should be addressed; however, very few developers give this the amount of attention it needs. For each domain, several criteria groups were collapsed, and criteria and benchmarks were combined to reflect SMEs feedback. We then validated the new criteria and benchmarks against industry standards, regulatory guidances, frameworks, and tools.

Finally, we conducted user testing with developers who were working on commercial products that fit ≥ 1 user group. We collected feedback from 9 developers, across the three domains, before reaching saturation. They indicated that the criteria and benchmarks were informative and applicable to their products, and suggested minor edits for evidence and privacy & security. The primary concern with usability was that the benchmarks were more detailed than those required by industry standards or regulatory bodies.

Domain: Evidence - Criteria and benchmarks covering the process of producing or gathering data, information, and findings through research, clinical trials, studies, or other methods	
Criteria group	Criteria
Evidence generation - Criteria covering the systematic approaches used to generate data demonstrating scientific validity and clinical evidence to support the product's claims	<p>Patient outcomes - Attestations covering measurable effects, results, or consequences experienced by patients as a result of intervention or treatment by the product</p> <ul style="list-style-type: none"> ● EG 1.1 - One or more studies have been performed that show the validity, reliability, and generalizability of patient outcomes driven by the software components of your product ● EG 1.2 - The study population is sufficiently diverse and/or representative of the patients for which the product intends to improve outcomes
	<p>Comparative effectiveness - Attestations covering the evaluation of how well the product performs relative to other options in real world settings</p> <ul style="list-style-type: none"> ● EG 2.1 - The product's safety has been compared to other leading products/standards of care in the industry ● EG 2.2 - The product's efficacy has been compared to other leading products/standards of care in the industry
Product impact - Criteria covering the effects, results, or consequences of use of the product on patients' health status, well-being, and quality of life	<p>End user experience - Attestations covering how users (such as healthcare professionals, patients, administrators, or caregivers) perceive and interact with the software</p> <ul style="list-style-type: none"> ● PI 1.1 - Potential risks associated with use of the product and their implications for the patient are clearly defined ● PI 1.2 - A process is in place for patients to report risks or unintended consequences associated with using the product ● PI 1.3 - Diverse stakeholders (e.g., clinicians, patients, caregivers) potentially affected by use of the product were engaged in its design, implementation, and evaluation ● PI 1.4 - Behavioral analytics, including but not limited to user behavior, system activity patterns, and network traffic are captured and analyzed to identify potential impacts to end user and/or patient outcomes
	<p>Cost measurement - Attestations covering the process of quantifying and assessing the expenses associated with the development, implementation, maintenance, and operation of the software</p> <ul style="list-style-type: none"> ● PI 2.1 - A cost analysis has been done for all paid products across a common baseline of direct and indirect costs
Domain: Privacy & Security - Criteria and benchmarks covering fundamental concepts that aim to protect users' data and ensure the confidentiality, integrity, and availability of software applications	
Criteria group	Criteria
Continuous monitoring - Criteria covering the ongoing process of real-time surveillance, analysis, and assessment of data to detect, prevent, and respond to threats, vulnerabilities, and incidents associated with the software	<p>Monitoring & analytics - Attestations covering practices and tools used to observe, measure, and analyze various aspects of software applications, systems, or infrastructure</p> <ul style="list-style-type: none"> ● CM 1.1 - Monitoring is implemented in production environments to identify indications of a privacy breach ● CM 1.2 - Behavioral analytics, including but not limited to user behavior, system activity patterns, and network traffic are captured and analyzed to detect anomalous or suspicious activities indicative of security threats
Data governance - Criteria covering processes of gathering, storing, using, and managing health-related information while ensuring confidentiality, integrity, and privacy are in place throughout the lifecycle of the data collected by the product	<p>Consent & authorization - Attestations covering voluntary agreements made by end users about how their data is collected, used, and distributed</p> <ul style="list-style-type: none"> ● DG 1.1 - Consent is being obtained and information is provided on the data being collected, how the data will be managed and shared (including anonymization or de-identification, use for marketing purposes), how long the data will be stored, why data collection is necessary for the use of the software, and what data will be monetized ● DG 1.2 - A mechanism is in place to make users aware of updates to privacy policies, terms of service, and/or other end-user agreements. Whenever updates are made to privacy policies, terms of service, and/or other end-user agreements
	<p>Data integrity & quality - Attestations concerned with the accuracy, reliability, and security of data as it is entered, stored, and retrieved. Additionally, the benchmarks evaluate if that data is relevant, accurate, and reliable</p> <ul style="list-style-type: none"> ● DG 2.1 - There is a process for reviewing data quality and quality assurance processes are in place to ensure the accuracy, completeness, and integrity of data collected
	<p>Anonymization & deidentification - Attestations covering techniques and methods that purposefully break the link between data values and individuals</p> <ul style="list-style-type: none"> ● DG 3.1 - Methods and reasons for anonymization, de-identification and re-identification are clear and adhere to HIPAA standards
	<p>Data retention - Attestations covering storage and maintenance of data for a specific period of time, determined by regulatory, legal, or business requirements</p> <ul style="list-style-type: none"> ● DG 4.1 - All reasons for health data retention are clearly explained and no additional purposes will be allowed unless explicit consent is asked from the patient ● DG 4.2 - Users are informed of their rights regarding the retention and disposal of their data, including the right to request access, correction, or deletion of their information
	<p>Data minimization - Attestations covering the specification of data elements necessary for providing the services associated with the use of the product</p> <ul style="list-style-type: none"> ● DG 5.1 - The product only collects data elements that are necessary for performing its intended function or are reasonably associated with product monitoring or development efforts that are aligned with its intended function
	<p>Access control - Attestations covering the process of selectively granting or denying access to data, systems, or physical locations based on the identity, role, or authorization level of the entity requesting</p> <ul style="list-style-type: none"> ● DG 6.1 - Authorization mechanisms are used to determine the permissions and privileges granted to authenticated users and third party vendors based on their roles, responsibilities, and access rights ● DG 6.2 - Auditing and logging capabilities track and record access attempts, user activities, and security events
	<p>Data encryption - Attestations covering processes and methods for ensuring sensitive information remains secure and unreadable to unauthorized users or attackers who might intercept the data</p> <ul style="list-style-type: none"> ● DG 7.1 - Encryption is applied to data stored on devices, servers, databases, and other storage systems to protect it from theft, loss, or unauthorized access to physical or digital storage ● DG 7.2 - Data encryption mechanisms provide secure data transmission over networks
Continued	

Domain: Evidence - Criteria and benchmarks covering the process of producing or gathering data, information, and findings through research, clinical trials, studies, or other methods	
Education & awareness - Criteria covering educational initiatives and programs designed to educate product developers and other stakeholders about best practices, policies, and procedures for data security in digital health environments	Security training - Attestations covering educational programs and initiatives aimed at educating software developers, engineers, IT professionals, and other personnel involved in software development and deployment about cybersecurity best practices, principles, and procedures ● EA 1.1 - Privacy and security training have been made available for all personnel with responsibilities that contribute to the development of your product, including but not limited to software and QA engineers, designers, product managers, and IT professionals
Health equity - Criteria covering the presentation of information for fair and equitable access and use of the software, regardless of socioeconomic status, academic attainment, age, or other factors	Consent - Attestations covering the permission or agreement given by a user for the collection, processing, and use of their personal data or for participating in certain activities facilitated by the product in a legally and ethically compliant manner ● HE 1.1 - Privacy policy and terms of service information, including the process of collecting consent, is written at an 8th grade reading level
Secure software development lifecycle - Criteria covering a systematic approach to integrating security practices and measures throughout the software development process	Development infrastructure - Attestations covering the underlying framework, tools, and resources used to create, test, and deploy your product ● SSDLC 1.1 - All forms of code – including source code, executable code, and configuration-as-code – are stored in a version controlled system based on the principle of least privilege so that only authorized personnel, tools, services, etc. have access
	Risk analysis - Attestations covering the process of identifying documenting, and validating the needs and expectations of stakeholders and the associated risks, as well as the strategies used to mitigate ● SSDLC 2.1 - A risk assessment was conducted for every system (database or API, including any 3rd party services) required for the functioning of the product. These risk assessments should include, for each system profiled: Descriptions of specific vulnerabilities, Types of risk to the organization of this vulnerability is exploited, Threat sources that could take advantage of this vulnerability, Existing safeguards that reduce this risk, Likelihood of occurrence, Ownership of monitoring and addressing this risk
	Third-party dependencies - Attestations covering external components, libraries, frameworks, modules, or services that are utilized within the software. ● SSDLC 3.1 - Dependency inventory, i.e., software bill of materials
	Integration testing & validation - Attestations covering any testing techniques used to verify the interactions between integrated components or modules of a software system ● SSDLC 4.1 - Integration tests are performed before major product releases to the product's end users
	Deployment & maintenance - Attestations covering the process of releasing the software application or system into a production environment and activities performed after deployment to ensure ongoing functionality, usability, and performance ● SSDLC 5.1 - Updates to security features are included in software release documentation to inform users about security considerations, where appropriate
Domain: Usability - Criteria and benchmarks covering how the product provides ways to easily and intuitively navigate interfaces, understand information, and perform tasks	
Criteria group	Criteria
User experience - Criteria covering how end users perceive and interact with your product, including overall experience, satisfaction, and ability to perform necessary software functions	Task efficiency - Attestations covering the effectiveness and speed with which users can accomplish specific tasks or goals within the product ● UE 1.1 - Navigational elements that appear on multiple screens generally follow the same order in each instance, unless: The user is able to customize their view, The context of the application has changed enough that the user would reasonably expect the navigation to change as well ● UE 1.2 - The product's content — such as text, images, video, and interactive elements — can be fully displayed without losing any information and without requiring the user to scroll both horizontally and vertically in the same view. This does not apply to content that requires both horizontal and vertical scrolling in order to maintain functionality, such as data tables or diagrams ● UE 1.3 - Elements that require user input within the product — such as form fields, buttons, and selectors — include labels that make the element's purpose clear. This includes labels that are pictographic in nature as long as the pictograph is clearly defined on the page (e.g., stars for required form fields) ● UE 1.4 - Elements that require user input within the product — such as form fields, buttons, and selectors — clearly indicate when a user input error is detected and the nature of the error is described in text ● UE 1.5 - Web pages and/or application screens include title elements that describe the page or screen so that users with visual, cognitive and motor disabilities, or limited short-term memory can determine where they are in the product ● UE 1.6 - Links and buttons within the product allow the user to determine the purpose of the element, including changes that might occur in the product once the link or button is pressed or locations the user might be redirected to
	Accessibility - Attestations covering the ways in which a product can be accessed and used by people with disabilities ● UE 2.1 - The product makes the appropriate semantic structure and elements available to the user agent to enable screen readers and other assistive devices ● UE 2.2 - All functionality is operable through a keyboard interface, enabling users with alternate keyboards or input devices to successfully use the product. Exceptions to this include any product functionality that is not compatible with keyboard input, such as drawing ● UE 2.3 - Text in the product can be resized without assistive technology, such as utilizing OS- or browser-provided zoom functionality. Exceptions to this include captions or images of text ● UE 2.4 - Information that is required for the product's core functionality is persisted after it is entered by the user, and is auto-populated or selectable when appropriate. This includes information such as credit cards, insurance information, and user info necessary for the functionality of the product
Continued	

Domain: Evidence - Criteria and benchmarks covering the process of producing or gathering data, information, and findings through research, clinical trials, studies, or other methods	
	User testing & support - Attestations covering processes and activities aimed at gathering feedback from users, evaluating their experiences, and providing assistance to ensure effective usage of the product <ul style="list-style-type: none">● UE 3.1 - A process is in place to provide the specific knowledge needed to effectively and proficiently use the product. Training is based on end-user role (e.g., patient, clinician, or administrator)● UE 3.2 - Usability testing is conducted with target users, representing each end-user role, to collect feedback on task completion and user satisfaction in conditions that match the product's intended use● UE 3.3 - A process is in place to document user feedback. This includes, but is not limited to, user feedback tools that are contained within the product or easily contactable customer support via phone, email, or similar
Workflow integration - Criteria covering the incorporation of a product into existing clinical and administrative workflows within healthcare settings	Workflow integration - Attestations covering the incorporation of a product into existing clinical and administrative workflows within healthcare settings <ul style="list-style-type: none">● WI 1.1 - Analysis of user roles, responsibilities, tasks, and decision points within the workflow align with product features and functionalities.● WI 1.2 - The product supports the appropriate FHIR endpoints so that it may interact with other systems, enhance patient care, and align with industry standards and regulations

Table 2. The evaluation framework with domains, criteria (groups) and associated benchmarks.

Discussion

The current disorganized processes for assessing the quality and trustworthiness of DHSPs can delay development and deployment of DHTs, and might not reflect best practices with regard to evidence, privacy and security, usability, and equity and inclusion. Though there are fragmented assessments focused on discrete domains of importance to evaluating DHSPs from the public and private sector, no industry-driven, non-partisan effort that unifies and provides market guidance to DHSPs and their adopters exists. An evidence-based EF would harmonize future efforts. This mixed-methods study created an EF to offer a comprehensive set of benchmarks for ensuring high quality of DHSPs.

Despite the existence of many regulatory guidances, frameworks, and industry standards, the SMEs we interviewed asked for additional specificity as to what “good” would look like for each domain we identified. Findings from the literature review echoed feedback from SMEs that an EF with clearly defined criteria and benchmarks is needed. Currently, the onus is on developers, adopters and end-users to identify relevant evaluation criteria and apply them. This is highly problematic: here are 6000 +³¹ hospitals in the US and 350,000 +¹ DHSPs for them to choose from with no standard approach to evaluation. The size of this systemic challenge is staggering and the implication to equitable care access is deeply concerning. Our framework provides comprehensive and clear parameters of quality and allows developers to attest their products against them so that adopters can more quickly identify and adopt fit-for-purpose DHSPs.

Many SMEs shared that very little consideration goes into evaluating usability for diverse populations and different end-users. This is consistent with research showing large inequities in access to and uptake in DHTs^{32,33}. Many developers stated that they rely on convenience sampling when conducting user testing. Similarly, developers shared that little effort is dedicated to accessibility, and both developers and adopters reported that they would like to see a bigger push to prioritize inclusive design in DHSP development.

The SMEs feedback also indicated a desire for greater transparency throughout the DHSP development and deployment phases. Diverse stakeholders should be included early in the DHSP development process^{10,34}. There are many opportunities for developers to be more transparent, including how evidence is generated to support the DHSP claims, providing terms and conditions for privacy and security that are easy for end-users to understand, and providing details on usability testing, among other aspects.

The EF is based on front-line voices and needs, and its modularity ensures it can be readily updated to stay current with the evolving needs of stakeholders. Its broad scope improves upon other evaluation tools that take a cost reduction^{35–37} or profit driven approach³⁸.

To retain the viability of this EF going forward we intend to maintain a transparent process upon which new benchmarks, standards, and criteria can be considered and incorporated. Our expectation for such evolutions are that they are (a) good for the DHSP adopters and developers, (b) enhance or maintain the relevance of this EF, and (c) are clear and attestable. This will allow for the current EF to consider important evolutions to this quickly changing landscape, including new guidelines developed after this effort was initially launched, such as new technological developments like AI and interoperability.

The study has a few limitations: Having greater sample sizes for interviews, focus groups, and surveys is often preferable but not always achievable. Most of the SMEs interrogated for the needs assessment represented senior executive or leadership roles, which may have skewed the results. We mitigated this by including mid-level participants in the cohort of SMEs that participated in the EF development phase.

Most of the benchmarks and guidelines leveraged to develop this EF are U.S.-focused. Though we included several standards with global or non-U.S. reach (e.g., GDPR), future extensions of the EF may more intentionally incorporate considerations related to other markets, e.g., such as the work happening on the European Digital Health Technology Assessment framework (EDiHTA)³⁹.

Finally, one limitation of inductive thematic analysis is that it relies on the researchers’ subjective assessments. We attempted to mitigate this by discussing the identified themes with DiMe internal experts to reach a consensus.

In conclusion, the proposed evidence-based EF spans multiple domains of trust and value, harmonizes best practices, stands on the shoulders of well-respected and established work, and eases the effective adoption of high-quality, trustworthy DHSPs. It also becomes a bedrock upon which future iterations can be built.

Data availability

All aggregated data generated as part of this study are included in this published article and its supplementary information files. The raw data collected as part of this research study (interview transcripts, surveys, literature reviews) are not publicly available and are available from the corresponding author on reasonable request.

Received: 2 April 2025; Accepted: 25 September 2025

Published online: 31 October 2025

References

1. IQVIA. Digital Health Trends. December 12, 2024. Accessed September 9, 2025. (2024). <https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/digital-health-trends-2024>
2. Gordon, W. J., Landman, A., Zhang, H. & Bates, D. W. Beyond validation: getting health apps into clinical practice. *NPJ Digit. Med.* **3**, 14. <https://doi.org/10.1038/s41746-019-0212-z> (2020).
3. Jacob, C. et al. Assessing the quality and impact of eHealth tools: systematic literature review and narrative synthesis. *JMIR Hum. Factors.* **10**, e45143. <https://doi.org/10.2196/45143> (2023).
4. Borges do Nascimento, I. J. et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. *NPJ Digit. Med.* **6**, 161. <https://doi.org/10.1038/s41746-023-00899-4> (2023).
5. Zakerbasali, S., Ayyoubzadeh, S. M., Baniasadi, T., Yazdani, A. & Abhari, S. Mobile health technology and healthcare providers: systemic barriers to adoption. *Healthc. Inf. Res.* **27** (4), 267–278. <https://doi.org/10.4258/hir.2021.27.4.267> (2021).
6. Rowland, S. P., Fitzgerald, J. E., Holme, T., Powell, J. & McGregor, A. What is the clinical value of mHealth for patients? *NPJ digit. Med.* **3**, 4. <https://doi.org/10.1038/s41746-019-0206-x> (2020).
7. Byambasuren, O., Beller, E. & Glasziou, P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. *JMIR Mhealth Uhealth.* **7** (6), e13199. <https://doi.org/10.2196/13199> (2019).
8. Dhruva, S. S. et al. Physicians' perspectives on FDA regulation of drugs and medical devices: A National survey. *Health Aff (Millwood).* **43** (1), 27–35. <https://doi.org/10.1377/hlthaff.2023.00466> (2024).
9. Torous, J., Stern, A. D. & Bourgeois, F. T. Regulatory considerations to keep Pace with innovation in digital health products. *NPJ Digit. Med.* **5**, 121. <https://doi.org/10.1038/s41746-022-00668-9> (2022).
10. Abernethy, A. et al. The promise of digital health: then, now, and the future. *NAM Perspect.* <https://doi.org/10.31478/202206e> (2022).
11. Marwaha, J. S. et al. Deploying digital health tools within large, complex health systems: key considerations for adoption and implementation. *NPJ Digit. Med.* **5**, 13. <https://doi.org/10.1038/s41746-022-00557-1> (2022).
12. RealTimeBoard Inc. dba Miro. Miro home page. Accessed August 6, 2024. (2024). <https://miro.com>
13. Zamanzadeh, V. et al. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. *J. Caring Sci.* **4** (2), 165–178. <https://doi.org/10.15171/jcs.2015.017> (2015).
14. Nikolopoulou, K. What is content validity? definition & examples. Scribbr. August 26, 2022. Accessed August 6, (2024). <https://www.scribbr.com/methodology/content-validity/>
15. Lawshe, C. A quantitative approach to content validity. *Person Psychol.* **38** (4), 563–575. (1975). <https://doi.org/10.1111/j.1744/6570.1975.tb01393.x>
16. Harzing, A. W. Publish or Perish. Accessed September 8, 2025. (2007). <https://harzing.com/resources/publish-or-perish>
17. Fereday, J. & Muir-Cochrane, E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *Int. J. Qual. Methods.* **5** (1), 80–92. <https://doi.org/10.1177/160940690600500107> (2006).
18. Thomas, J. & Harden, A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med. Res. Methodol.* **8**, 45. <https://doi.org/10.1186/1471-2288-8-45> (2008).
19. World Health Organization. Performing a landscape analysis: a quick guide. August 29. Accessed August 6, 2024. (2023). <https://www.who.int/publications/i/item/9789240073319>
20. HITRUST Services Corp. The global standard of information protection assurance: HITRUST. Accessed August 7, 2024. (2024). <https://hitrustalliance.net>
21. AICPA Assurance Services Executive Committee. SOC 2[®] description criteria (with revised implementation guidance – 2022). 2022. Accessed August 7, 2024 (2018). <https://www.aicpa-cima.com/resources/download/get-description-criteria-for-your-organizations-soc-2-r-report>
22. Health Insurance Portability and Accountability Act of 1996; (1996). <https://www.congress.gov/104/plaws/publ191/PLAW-104publ191.pdf>
23. US Department of Health and Human Services, Office for Civil Rights. HITECH Act Enforcement Interim Final Rule. October 28. Accessed August 7, 2024. (2009). <https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html>
24. SMART Health IT. SMART on FHIR. Accessed August 7, 2024. (2020). <https://docs.smarthealthit.org/>
25. US Food and Drug Administration, Center for Devices and Radiological Health. Software as a medical device (SaMD). FDA. September 9, 2020. Accessed August 7. (2024). <https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd>
26. National Committee for Quality Assurance (NCQA). HEDIS measures and technical resources. NCQA. Accessed August 7, 2024. (2024). <https://www.ncqa.org/hedis/measures/>
27. Military Health System. Joint Health Information Exchange. Military Health System. April 23. Accessed August 7, 2024. (2024). <https://health.mil/Military-Health-Topics/Technology/Joint-HIE>
28. Office of the National Coordinator for Health IT. About the ONC health IT certification ehrs/about-onc-health-it-certification-program. November 9. Accessed August 7, 2024. (2021). <https://www.healthit.gov/topic/certification-ehrs/about-onc-health-it-certification-program>
29. KLAS Research. Home page. Accessed August 7. (2024). <https://klasresearch.com/>
30. European Union. General Data Protection Regulation (GDPR). General Data Protection Regulation (GDPR). May 25, 2018. Accessed August 7. (2024). <https://gdpr-info.eu/>
31. American Hospital Association. Fast Facts on & Hospitals, U. S. (2025). <https://www.aha.org/statistics/fast-facts-us-hospitals>
32. Woolley, K. E. et al. Mapping inequities in digital health technology within the world health organization's European region using PROGRESS PLUS: scoping review. *J. Med. Internet Res.* **25**, e44181. <https://doi.org/10.2196/44181> (2023).
33. Wilson, S. et al. Recommendations to advance digital health equity: a systematic review of qualitative studies. *NPJ Digit. Med.* **7** (1), 1–9. <https://doi.org/10.1038/s41746-024-01177-7> (2024).
34. Largent, E. A., Karlawish, J. & Wexler, A. From an Idea to the marketplace: identifying and addressing ethical and regulatory considerations across the digital health product-development lifecycle. *BMC Digit. Health.* **2** (1), 41 <https://doi.org/10.1186/s44247-024-00098-5> (2024).
35. Neumann, P. J., Willke, R. J. & Garrison, L. P. A health economics approach to US value assessment frameworks-introduction: an ISPOR special task force report. *Value Health.* **21** (2), 119–123. <https://doi.org/10.1016/j.jval.2017.12.012> (2018).
36. Baldwin, C. & von Hippel, E. Modeling a paradigm shift: from producer innovation to user and open collaborative innovation. *Organiz. Sci.* **22** (6), 1399–1417. <https://doi.org/10.1287/orsc.1100.0618> (2011).

37. Berwick, D. M., Nolan, T. W. & Whittington, J. The triple aim: care, health, and cost. *Health Aff (Millwood)*. **27** (3), 759–769. <https://doi.org/10.1377/hlthaff.27.3.759> (2008).
38. Hwang, J. & Christensen, C. M. Disruptive innovation in health care delivery: a framework for business-model innovation. *Health Aff (Millwood)*. **27** (5), 1329–1335. <https://doi.org/10.1377/hlthaff.27.5.1329> (2008).
39. EDiHTA Project. The first European Digital Health Technology Assessment framework co-created by all stakeholders in the European Health Ecosystem. Accessed September 9. (2025). <https://edihta-project.eu/>

Acknowledgements

The authors thank Jennifer Ostertag-Stretch MBA (DiMe Society, Boston MA) for her insights and assistance in development of the framework, and Rachell Chon, Sukhrob Makhkamov, John Scott Wrigley, and Anya Odabasic for assistance with the literature review.

Author contributions

JG conceptualized the work and research methodology, and edited the original draft. YS and CH conducted data collection and result analysis. DM and CZ helped to conceptualize and validate the work, and reviewed the original draft. BV supervised the data analysis and wrote the original draft. All authors read and approved the final manuscript.

Declarations

Conflict of interest

DiMe built the DiMe Seal, a web-based platform that implements the evaluation framework discussed in this manuscript. Our intent to publish is to make the underlying approach open access so that others can vet our work and build from it.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-21996-2>.

Correspondence and requests for materials should be addressed to B.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025