Methods, Indicators, and End-User Involvement in the Evaluation of Digital Health Interventions for Public: A Scoping Review

Vera Weirauch, Clarissa Soehnchen, Anja Burmann, Sven Meister

Submitted to: Journal of Medical Internet Research
on: December 22, 2023

Table of Contents

Methods, Indicators, and End-User Involvement in the Evaluation of Digital Health Interventions for Public: A Scoping Review

Vera Weirauch[1, 2] MA; Clarissa Soehnchen[1] MLB; Anja Burmann[2] Dr rer med; Sven Meister[1, 2] Prof Dr

[1]Health Informatics Faculty of Health, School of Medicine Witten/Herdecke University Witten DE
[2]Fraunhofer Institute for Software and Systems Engineering Dortmund DE

Corresponding Author:
Vera Weirauch MA
Fraunhofer Institute for Software and Systems Engineering
Speicherstraße 6
Dortmund
DE

Abstract

Background: Digital health interventions (DHIs) have the potential to enable public end-users, such as citizens and patients, to manage and improve their health. Although the number of available DHIs is increasing, examples of successfully established DHIs in public health systems are limited. To counteract a non-use of DHIs they should be comprehensively evaluated while integrating end-users. Unfortunately, there is a wide variability and heterogeneity according to the ways of evaluation which creates a methodological challenge.

Objective: The scoping review aims to provide an overview of the current established processes for evaluating DHIs, including methods, indicators and end-user involvement. The review is not limited to a specific medical field or type of DHI but offers a holistic overview.

Methods: This review was conducted based on the methodology framework for scoping reviews of Arksey & O'Malley and complies with the PRISMA-ScR guideline. Three scientific databases (Pubmed, Scopus, Science Direct) and grey literature were searched in April 2023. English and German studies between 2008 and 2023 were considered when a DHI that explicitly addresses public end-users was evaluated.

Results: The search strategy identified 9618 publications, of which 160 were included. Of these finally included, 200 evaluations were derived and analysed. It is discovered that there is neither a consensus on methods to evaluate DHIs nor a commonly agreed definition or usage of the evaluated indicators which results in a broad variety of evaluation practices. This aligns with observations of existing literature. As assistance for people developing and evaluating DHIs and as a basis for thinking about appropriate ways to evaluate DHIs, a result matrix is created where the findings were combined per DHI cluster. Additionally, general recommendations for evaluators of DHIs are formulated.

Conclusions: The findings of this scoping review offer a holistic overview of the variety and heterogeneity according to the ways DHIs for public end-users are evaluated. Evaluators of these DHIs should be encouraged to reference established frameworks or measurements for justification. This would ease the transferability of results among similar evaluation studies within the digital health sector, thereby enhancing the coherence and comparability of research in this area.

(JMIR Preprints 22/12/2023:55714)
DOI: https://doi.org/10.2196/preprints.55714

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
    Please make my preprint PDF available to anyone at any time (recommended).
    Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
    Only make the preprint title and abstract visible.
✓ **No, I do not wish to publish my submitted manuscript as a preprint.**
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Review

Methods, Indicators, and End-User Involvement in the Evaluation of Digital Health Interventions for Public: A Scoping Review

Vera Weirauch[12], Clarissa Soehnchen[1], Anja Burman[2], Sven Meister[12]
[1]Health Informatics, Faculty of Health/School of Medicine, Witten/Herdecke University, Witten, Germany
[2]Fraunhofer Institute for Software and Systems Engineering ISST, Dortmund, Germany

Abstract

Background: Digital health interventions (DHIs) have the potential to enable public end-users, such as citizens and patients, to manage and improve their health. Although the number of available DHIs is increasing, examples of successfully established DHIs in public health systems are limited. To counteract a non-use of DHIs they should be comprehensively evaluated while integrating end-users. Unfortunately, there is a wide variability and heterogeneity according to the ways of evaluation which creates a methodological challenge.

Objective: The scoping review aims to provide an overview of the current established processes for evaluating DHIs, including methods, indicators, and end-user involvement. The review is not limited to a specific medical field or type of DHI but offers a holistic overview.

Methods: This scoping review was conducted following the JBI methodology for scoping reviews based on Arksey & O'Malley and complies with the PRISMA-ScR guideline. Three scientific databases (Pubmed, Scopus, Science Direct) were searched in April 2023. English and German studies between 2008 and 2023 were considered when a DHI that explicitly addresses public end-users was evaluated. The process of study selection was carried out by several researchers to avoid reviewer bias.

Results: The search strategy identified 9618 publications, of which 160 were included. Of these finally included, 200 evaluations were derived and analysed. The results show that there is neither a consensus on methods to evaluate DHIs nor a commonly agreed definition or usage of the evaluated indicators which results in a broad variety of evaluation practices. This aligns with observations of existing literature. It is discovered that there is a lack of references to existing frameworks for the evaluation of DHIs. The majority of the included studies referred to user-centred approaches and involved end-users in the evaluation process. As assistance for people developing and evaluating DHIs and as a basis for thinking about appropriate ways to evaluate DHIs, a result matrix is created where the findings were combined per DHI cluster. Additionally, general recommendations for evaluators of DHIs are formulated.

Conclusions: The findings of this scoping review offer a holistic overview of the variety and heterogeneity according to the ways DHIs for public end-users are evaluated. Evaluators of these DHIs should be encouraged to reference established frameworks or measurements for justification. This would ease the transferability of results among similar evaluation studies within the digital health sector, thereby enhancing the coherence and comparability of research in this area.

Keywords: digital health; digital health intervention; public end-user; evaluation methods; evaluation criteria; end-user involvement; scoping review

Introduction

Background

The potential of digital health interventions (DHIs) to improve care processes is widely recognized [1,2]. Particularly for public end-users, DHIs have the potential to enable them to manage and improve their health by promoting health, supporting behaviour change, personalizing healthcare delivery, and giving them the chance to individually organize healthcare [1–6]. Consequently, the development of digital health interventions is increasing, so that a growing number of them are available in the market [3,4]. Nevertheless, examples of successfully established DHIs in public health systems are limited [7] and several challenges exist in this rapidly growing field, for example, methodological challenges [1,6,8,9].

Before elaborating on the challenges, it is useful to consider the term DHI from a definitional point of view. There are various terms for digital technologies in the health sector, such as eHealth, mHealth, digital health services, and digital health intervention. To avoid conceptual ambiguity, this paper adopts the World Health Organisation’s (WHO) definitional approach, defining a digital health intervention as “a discrete functionality of the digital technology to achieve health sector objectives”[10]. Following this definition, the growing field of digital public health interventions (DPHIs) can be regarded as a distinct subset of DHIs. DPHIs primarily focus on improving health and well-being at the population level, rather than at the individual level [11–13]. For this study, the definition of key term evaluation is also adopted by the WHO, defining evaluation as “The systematic and objective assessment of an ongoing or completed intervention, with the aim of determining the fulfilment of objectives, efficiency, effectiveness, impact and sustainability”[14]. Additionally, it is important to specify that in the context of this scoping review, 'public end-users' are defined as individuals, such as citizens and patients, who directly interact with digital health tools or services. Unlike healthcare professionals or caregivers, who may use DHIs as part of their job, public end-users engage with these tools to meet their personal health needs. They are not limited to any specific patient group or demographic.

Status quo on evaluation of DHIs

Overall, DHIs are characterised as complex interventions. This complexity is partly due to their interdisciplinary and multisectoral nature, involving a diverse mix of stakeholders, including patients, various health professionals, relatives, policymakers, and health insurers [1,3,8,15]. Additionally, DHIs consist of multiple interdependent components, both technical and non-technical [4,8,15] and serve multiple aims, such as providing information, improving communication, facilitating data sharing, and enabling monitoring [1]. The value and impact of DHIs extend beyond clinical outcomes and also encompass organisational, behavioural and technical dimensions [7]. These characteristics also apply to DPHIs [11–13]. Consequently, the evaluation of DHIs as well as DPHIs is a methodological challenge [1,6,8,9] and existing, established methods like health technology assessment (HTA), evidence-based medicine (EBM), and randomised controlled trials (RCTs), are limited in their application [16–18]. Health Technology Assessment (HTA) is a process for the systematic evaluation of medical procedures and technologies, with a focus on assessing the health benefits and costs associated with the use of therapeutics, medical products, and diagnostic procedures [17]. Therefore this methodology is not immediately appropriate in the context of the evaluation of DHIs [17,19]. Evidence-based medicine (EBM) means that decisions about the care of patients should be based on the best available external clinical evidence from systematic research in combination with the clinical expertise of the health professional. Regarding DHIs, there are more aspects to evaluate than clinical aspects, so it requires other distinct approaches than the usual suggested process to gather evidence within EBM [20]. Although randomised controlled trials (RCT) are considered as the gold standard [6], especially in context of evaluating DHIs they are a muchdiscussed topic. Due to the complexity of DHIs, the applicability of RCTs is widely criticised [3,6,9,15,16,18]. A comprehensive evaluation of DHIs is essential for generating robust evidence [14,20], contributing to their long-term successful implementation and aiding in the realisation of the full benefits of DHIs [15,20]. This principle also applies to DPHIs, which should adhere to an evidence- and needs-based approach, incorporating a participatory, user-targeted development design to enhance acceptance of the intervention within the population [11,13,21].

The literature describes considerable variability and heterogeneity in how DHIs should be evaluated attributable to the absence of a standardised, established, and broadly applicable approach for evaluating DHIs [3,6,15,20,22–25]. One possible explanation for the absence of a standardised and broadly applicable guideline for the evaluation of DHIs is that the evaluation of these technologies is complex and complicated by various fundamental issues [17]. Various European authorities and scientists have addressed this issue by developing and publishing proposed frameworks for the evaluation of DHIs such as the Monitoring and Evaluation Guideline of the WHO, the Evidence standards framework for digital health technologies of the National Institute for Health and Care Excellence (NICE), the Report of the Expert Panel on effective ways of investing in Health (EXPH), the Swiss Evaluation framework by Kowatsch et al.[24], and the approach of Murray et al. [1,5,14,17,24,26]. In this review, there is an orientation towards the WHO and NICE frameworks, for instance, in categorising the evaluated DHIs and determining the evaluation criteria.

Status quo of end-user involvement in evaluation of DHIs

For various reasons, such as counteracting the non-use of DHIs as well as DPHIs, evidence generation and evaluation should be practice-oriented, necessitating the integration of end-users in this process [3,11,13,15,21,25,27–29]. In the recent years, approaches such as user-centred design (UCD), participatory health research (PHR) and public and patient involvement (PPI), have gained increasing importance in the healthcare sector, each contributing to the overarching goal of creating patient-centred, accessible, and equitable healthcare solutions.

UCD, or human-centred design, is rooted in human-system interactions and can be seen as a set of principles and strategies in the design and development of interactive digital health solutions, emphasizing the iterative research, design, and evaluation of services and systems by involving end-users and stakeholders throughout the project life cycle [30–32]. PHR can be understood as a research paradigm rather than a research method, aiming to increase the participation of individuals whose lives or work are the subject of research throughout the entire research process. The research process should be realised as a partnership among involved stakeholders, meaning with each other, instead of making research on people as passive objects. Involved stakeholder could include academic researchers, health professionals, policymakers, and members of civil society [33]. INVOLVE, founded by the National Institute for Health Research (NIHR) and taken over by the NIHR Centre for Engagement and Dissemination in April 2020, defines PPI as research that is realised ‘with’ or ‘by’ members of the public, rather than ‘to’, ‘about’ or ‘for’ them [34]. The term ‘public and patients’ includes current, former as well as potential patients, people who use health and social care services as well as people from organizations who represent other people using these services, and carers. [35–37].

To sum up, regarding the definitional approaches of PHR and PPI, overlaps can be seen, especially regarding the statements that research should be done with the research subjects rather than about them. UCD aligns with this principle but is more specifically focused on the design, development and evaluation of interactive digital health solutions, than on the scientific research context.

Through the involvement of end-users in evaluation processes, acceptance and usability problems can be mitigated [15,27] and health interventions can be designed in a target group-specific and needs-based manner [28].

Objectives

This scoping review is conducted to understand and provide an overview of the current established processes for evaluating digital health interventions for public end-users, such as citizens or patients. Previous reviews have focused on investigating which aspects of DHIs were evaluated during different development phases [20], investigating evaluation methods regarding specific criteria [29,38,39] or systems [40], investigating evaluation methods in specific medical contexts [22], investigating concrete evaluation methods despite RCTs [16], investigating economic evaluation of preventive digital public health interventions [12], or investigating general methods to evaluate effects of DHIs for citizens by making a review about reviews [6]. This review differs from those mentioned earlier as it provides a holistic overview of evaluation processes for DHIs for public end-users, not limited to a specific medical field. The objectives are to (1) capture the evaluation methods and (2) evaluation criteria that are currently used to evaluate DHIs. Additionally, there is a focus on (3) investigating the involvement of public end-users, in the (further)development of DHIs specifically developed for them. Therefore, the review addresses the following research questions (RQs):

1. Which research methods are used to evaluate digital health intervention for individuals/ public end-users?
2. Which evaluation criteria can be identified? Which evaluation criteria have been investigated?
3. In which way are individuals/ public end-users involved in the evaluation process?

To summarise, this scoping review aims to provide an overview of the currently established processes for evaluating DHIs, including methods, indicators, and end-user involvement.

Methods

Through a scoping review, broad topics can be explored and gaps in the evidence can be identified [41,42]. Due to the nature of scoping reviews, we did not formally assess the risk of bias or methodological quality of the included studies [41,42]. This review was conducted based on the methodology framework for scoping reviews of Arksey & O'Malley [43] and adheres to the PRISMA-ScR guideline (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) [44] incorporating the updates published by Peters et al. [45], see Multimedia Appendix 1, Table A1. The review protocol was registered a priori with the Center for Open Science (OSF) [46].

Search Strategy

The primary information sources for this scoping review are scientific databases, namely PubMed, Scopus, and ScienceDirect. Additionally, Google Scholar and the reference lists of included papers were manually screened. The search was conducted in April 2023. The search string, developed iteratively by three domain experts, was equally applied across the formerly mentioned databases, and is detailed in Table 1. The search was carried out without the assistance of librarians. The main search terms—'evaluation', 'digital health intervention', and 'user-centred'—along with their synonyms, were combined using Boolean operators.

Table 1- Search string

AND	**Field: Evaluation (A)**	**Context: DHI (B)**	**Focus: (C)**
OR	evaluat*, evaluation	digital health intervention*, digital health technology,	User-oriented, User-centered

	method* formative evaluation, summative evaluation, assess*	digital public health intervention*, digital health service*, electronic health record, mHealth, eHealth, health information technology, health Information platform, health diary,

The search string for each of the three databases are listed in Multimedia Appendix 1, Table A2. The results of each database were stored in Citavi and exported into Excel files.

Eligibility Criteria

Literature is eligible for inclusion if it describes evaluation methods or evaluation criteria for DHIs primarily aimed at public end-users. Additionally, the DHIs must be usable without the assistance of a health professional. The scoping review includes all types of DHI, including DPHIs without any exclusions, as long as the evaluation included an explorable version of the DHI. Moreover, literature that documents the integration of end-users in the evaluation process is also eligible for inclusion. According to a previous manual search, it appears that the term DHI was more commonly used approximately 15 years ago. Additionally, the authors assumed that the introduction of the iPhone in 2007 led to an increase in mHealth evaluation studies. Therefore, literature published in the last 15 years (2008-2023) was included in the search. The eligibility criteria relevant to this scoping review are detailed in Table 2.

Table 2- Eligibility criteria

	Inclusion Criteria	**Exclusion Criteria**
Targeted Population	Primary end-users of DHIs are individuals, such as patients or public. The DHI can be used on their own. No limitations on the number of participants, their gender, or their origin.	Primarily end-users are health professionals. The DHI can only be used with assistance of a health professional.
Study design	Original peer-reviewed studies, conference papers, book chapters, and grey literature such as organisational reports.	Not peer-reviewed papers, preprints, reviews, comments, presentations, protocols, or posters.
Context-Field DHI	DHIs can occur, for example, in form of patient portals, platforms, web or mobile applications or patients' access to electronic health records. No exclusion of specific types of DHIs. There should be an explorable version of the DHI, which means at least a low-fidelity prototype, within the evaluation.	There is no existing explorable version of the DHI.

Context-Field Evaluation	The study proposes or describes the evaluation/assessment of a DHI. Iterative evaluations primarily used for requirement engineering or assessing basic needs are excluded/ not focused.	The paper describes a framework or a general overview of evaluation types and methods. The study only focuses on RE processes, which means there is no explorable DHI (prototype or final DHI) available. Studies addressing only specific health issues designed to answer clinical research questions.
Context-Field End-User involvement	End-users are kind of actively integrated in the evaluation process.	The study did not include end-user in the evaluation process. End-users are only passive data objects, which means that they are not addressed with specific questions.
Accessibility	Full text is freely available on the internet or after contacting the author	Full text is not available on the internet or through contacting the author
Language	English, German	Others than English and German
Year	Published in the last 15 years (2008-2023)	Literature that is older than 15 years

Process of study selection

The electronic search results were stored in Microsoft Excel. An initial selection based on language and publication year was partially conducted within the scientific databases. To select the search results firstly duplications were removed. Subsequently, titles and abstracts were screened, and literature not meeting the eligibility criteria was excluded. To avoid reviewer bias, the screening process was partially conducted by different researchers. In concrete, a random sample of 400 titles (400/2896, 13.8%) was additionally screened by two independent researchers. Furthermore, the entire abstract screening was conducted by two independent researchers. In case of ambiguities regarding eligibility, discrepancies were discussed until a consensus was reached. Finally, full texts were screened against the eligibility criteria.

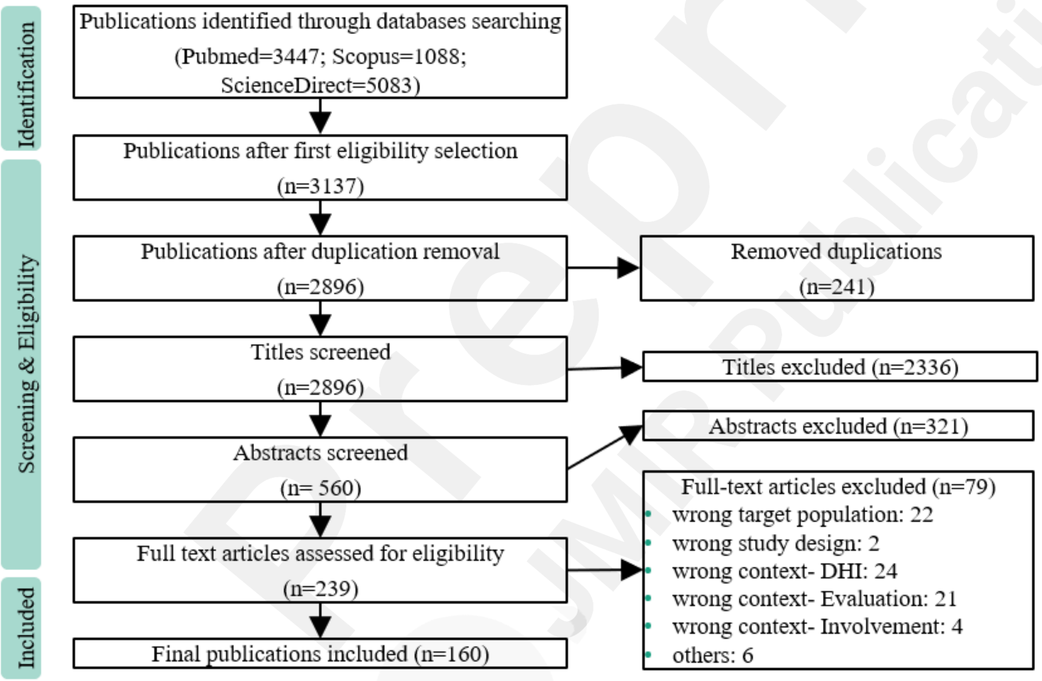
Data Extraction and Analysis

The data extraction was conducted to categorise the included papers for evidence synthesis. Data from the included sources were systematically extracted and organised into a pre-developed Excel spreadsheet. The data coding sheet was created within the research team and refined iteratively. The following data were extracted: bibliographic information (e.g., author, year of publication, title, DOI), characteristics of evaluated DHI (e.g., type of DHI- classified regarding WHO & NICE, addressed medical issue, intended use setting), evaluation methods (e.g. study design, usage of standardised approaches, amount of different methods, each method such as surveys/questionnaires, interviews, task-/scenario-completion, thinking-aloud, system data analysis, free testing phase including their duration, focus groups and others), evaluation criteria (e.g., aspect of assessment, explanation or definition of evaluated criteria or indicators, evaluated indicators such as clinical outcomes, user behaviour change, user experience, technical performance, content performance, actual system usage, suggestions for improvement and others), type of end-user involvement (passive data object, active data object, qualitative data subject). As part of this phase, classification schemes for the DHIs and the evaluated indicators were formulated using inductive category formation, whereby the extracted data served as the basis. This aimed to cluster all extracted DHIs and indicators within the context of this scoping review. The key terms used were outlined in Multimedia Appendix 1, Tables A3 and A4, considering the wide array of DHIs as well as evaluated indicators along with the lack of a universally used classification framework.

Results

A total of 9618 articles were identified in April 2023 from the search strategy across the three scientific databases utilised. Through an initial selection in the databases and the removal of duplicates, 2896 records remained for title screening. After this phase, 560 records remain for abstract screening. A number of 239 records were selected for full-text screening. During this stage, 79 records were excluded. Finally, 160 full texts met the eligibility criteria, resulting in 200 evaluations being derived and analysed. This is explained by the observation that several papers (38/160, 23,8%) included multiple rounds of evaluation, each employing different methods. For a brief overview of the included paper, the number 160 is relevant, whereas for analysis purposes, the number of evaluations resulting from the included papers is relevant. The screening process is shown in Figure 1.

Figure 1- PRISMA workflow chart



Characteristics of included studies and DHIs examined in these

The included studies were published between 2010 and 2023, whereas the number of published studies grows with the year. Specifically, 59 studies (59/160, 36.9%) were published before 2020, and 101 studies (101/160, 63.1%) from 2020 to April 2023.

In order to categorize the DHIs investigated in the paper, it was intended to use the WHO and NICE classification schemes [5,10]. There were many different kinds of DHIs apparent, ranging from simple websites for information purposes to complex multimodal interventions. However, these established classification schemes reach their limits, as DHIs often serve multiple functions. As a result, more than one-quarter of the DHIs (44/160, 27.5%) could not be clearly categorised within either scheme. Therefore, a summarising classification scheme to be used in the context of the scoping review was formulated by using inductive category formation whereby the extracted data served as the basis (see Figure 3 and Multimedia Appendix 1, Table A2.). According to this scheme, most DHIs (42/160, 23.8%) support the self-management of health and care followed by DHIs that are used as a digital supportive component of a treatment (31/160, 19.4%).

In 66 papers (66/160, 41%), the primary focus was on the evaluation process. In 94 papers (94/160, 59%), the evaluation process was addressed secondarily, as part of a comprehensive description of the DHI development. Table 3 summarises the characteristics of the included studies and the DHIs examined in these.

Table 3- Characteristics of included paper (N=160)

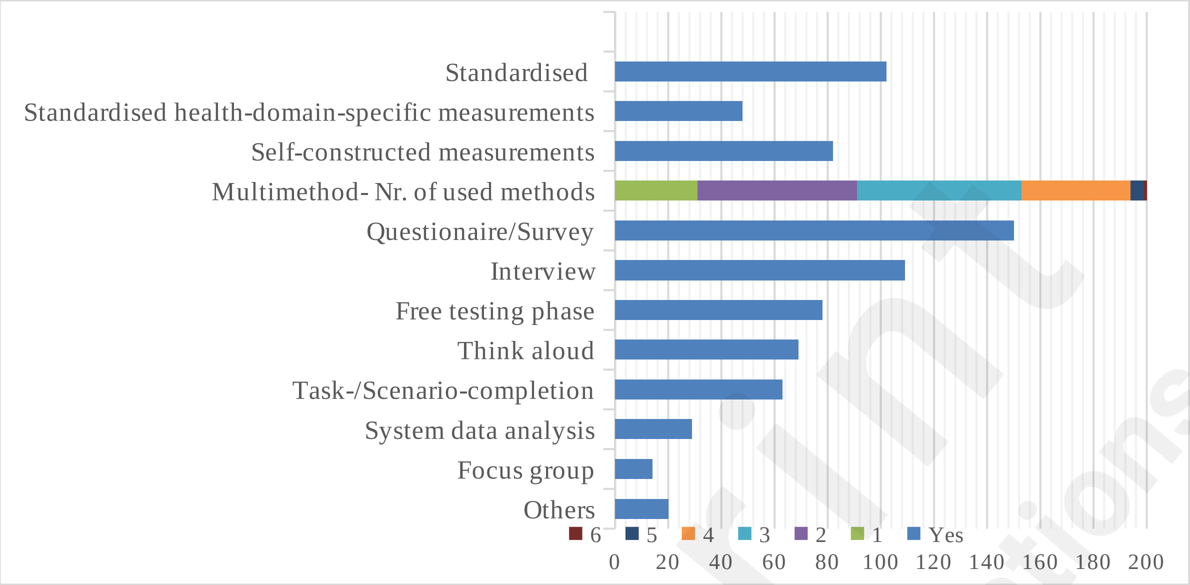
Characteristics		Value
Year of publication, n (%)		
	2010-2011	2 (1.3)
	2012-2013	4 (2.5)
	2012-2013	5 (3.1)
	2012-2013	17 (10.6)
	2012-2013	31 (19.4)
	2012-2013	52 (32.5)
	2012-2013	49 (30.6)
Focus of Paper, n (%)		
	Primarily evaluation	66 (41)
	Secondary evaluation	94 (59)
DHI categorisation in established schemes, n (%)		
	Unclear or multiple	44 (27.5)
	Categorised	116 (72.5)
DHI categorization in newly developed scheme, n (%)		
	Interaction with care provider- data transfer	2 (1.3)
	Interaction with care provider- communication	5 (3.1)
	Monitoring	22 (13.7)
	Tailored Information	22 (13.7)
	Non-tailored information	24 (15)
	Digital supportive treatment component	31 (19.4)
	Self-management	54 (33.8)
Addressed medical issue of DHI, n (%)		
	Prevention/ Promotion	6 (3.8)
	Generic	20 (12.5)
	Mental	33 (20.6)
	Somatic	101 (63.1)
Intended use setting of DHI, n (%)		
	Nursing/ Retirement home setting	1 (0.6)
	Rehabilitation setting	7 (4.4)
	Clinical/ stationary setting	8 (5)
	Ambulant/ primary care setting	40 (25)
	Prevention/ Promotion	42 (26.2)
	General health care setting	62 (38.8)

Overview: Evaluation Methods

To answer RQ1, “Which research methods are used to evaluate digital health intervention for individuals/ public end-users?”, the extracted data were analysed to provide an overview of the used methods.

Analysis of the 200 evaluations revealed that a mixed-methods study design was most commonly used (121/200, 60.5%), followed by qualitative (42/200, 21%) and quantitative (37/200, 18.5%) study designs. Mostly, a combination of 2 (60/200, 30%) or 3 (62/200, 31%) different research methods were applied. Surveys were the most commonly used method (150/200, 75%), followed by interviews (109/200, 54.5%) and testing phases (78/200, 39%). The duration of the testing phases predominantly exceeded one month (31/200, 15.5%), followed by durations between 1 week and 1 month (21/200, 10.5%), and those shorter than one day (13/200, 6.5%). Figure 2 provides a detailed overview of specific research methods used.

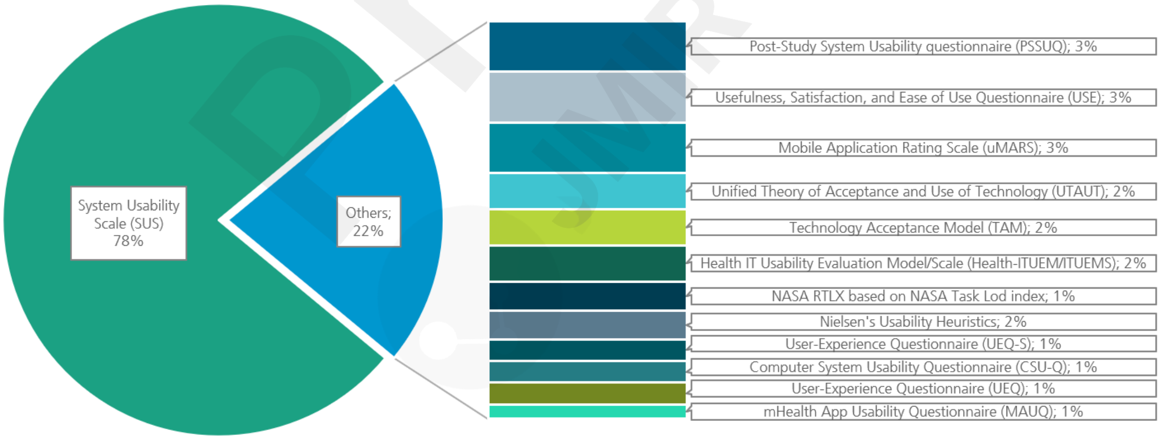
Figure 2- Research methods



As depicted in Figure 2, in over half of the evaluations (102/200, 51%), standardised usability scales, measurements, or questionnaires were utilised. Self-constructed measurements were used in 82 evaluations (82/200, 41%), while standardised health-domain-specific measurements were used in 48 evaluations (48/200, 24%).

To examine the commonly used standardised "usability" frameworks, Figure 3 visualises most of the used measurements. For measurements that appeared in fewer than two evaluations, only health-specific ones were listed.

Figure 3- Overview of standardised usability measurements



Overview: Evaluation Criteria

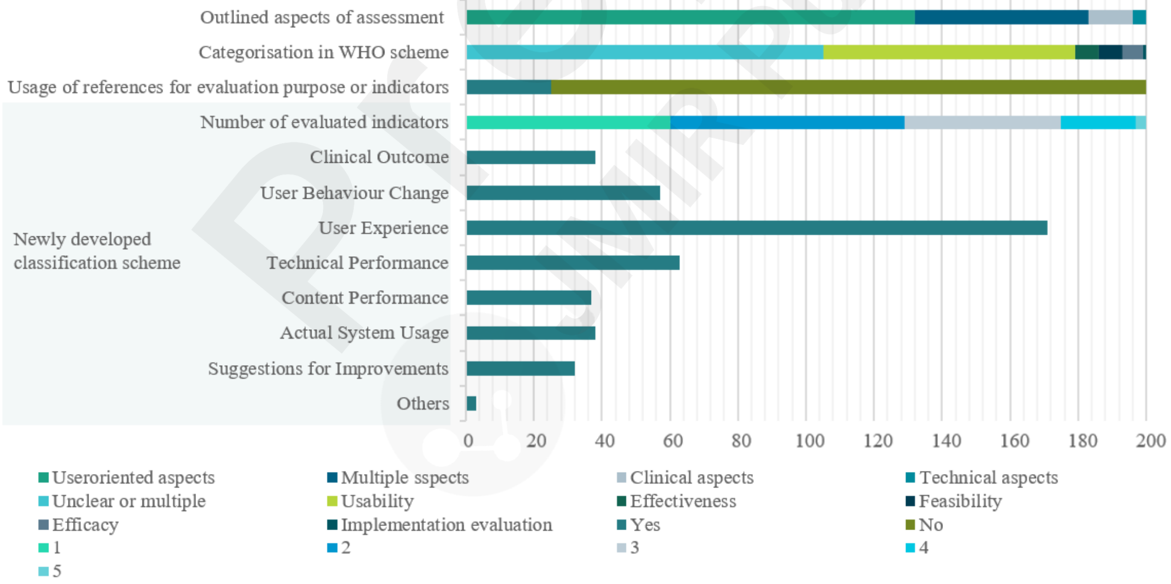
In order to answer RQ2 "Which evaluation criteria can be identified? Which evaluation criteria have been investigated?" extracted data were analysed to create a holistic picture of the used evaluation criteria.

Two-thirds of the generally outlined aspects of assessment are user-oriented (132/200, 66%), followed by multiple aspects (51/200, 25.5%), clinical outcomes (13/200, 6.5%), and technical aspects (4/200, 2%). In order to categorise the evaluation criteria, it was intended to map the extracted evaluations into the established classification scheme of the WHO, which contains feasibility, usability, efficacy, effectiveness, and implementation research [14]. Similarly, when attempting to map evaluation criteria into the WHO classification scheme, limitations were encountered as there are multiple criteria mentioned by the authors or a classification is unclear because the terms are used differently. In concrete, the underlying descriptions or definitions of the terms based on a variety of approaches and sources and additionally the used terms also vary. For example, in some evaluations usability is defined and therefore measured by underlying indicators like effectiveness, efficiency, satisfaction [47–50] or others like acceptance or feasibility [51,52]. In other cases, some of the previously listed underlying indicators are either indicators of other criteria like feasibility [53–56] or are seen as independent criteria, as for example by the WHO or others [14,57–61]. Consequently, more than half of the evaluations (105/200, 52.5%) could not be clearly mapped to one of the WHO-described criteria, as shown in Figure 4.

To further explore the term usage, the sources, or references to which the authors of the included paper refer were examined. It was apparent that in 87,5% of the evaluations (175/200, 87.5%), no information or literature references were provided to guide the terms used. Out of the 25 evaluations (25/200, 12.5%) that referred literature to describe the used terms, mostly ISO 9241-11 is referred (12/200, 6%), followed by multiple authors (9/200, 5%), Nielsen (3/200, 1.5%) and other references used once like Hix & Hartson or FITT framework.

As an alternative to the WHO classification scheme, a new scheme was formulated for use in this scoping review. Inductive category formation was used to form the new scheme, whereby all indicators extracted from the evaluations served as the data basis. This process led to the consolidation of eight criteria dimensions, described in Multimedia Appendix 1, Table A4, along with their associated indicators. Figure 4 visualises the described findings and the classification according to the newly developed scheme.

Figure 4- Research criteria



Overview: End-User-Involvement

In order to answer RQ3 “In which way are individuals/ public end-user involved in the evaluation process?” extracted data were analysed to investigate the end user involvement in the evaluation of DHIs that are primarily used by them.

To investigate end-user involvement, the ways they were involved have been categorised into (1) passive data objects, where citizens or patients do not have an active role, such as in responding to surveys and evidence is gathered, for example, by analysing system data; (2) active data objects, characterised by consciously responding to surveys with predefined options or participating in task/scenario-based sessions; and (3) qualitative data objects, where patients or citizens provide individual responses, allowing them to express and explain their views and emotions. These groups are further described in Multimedia Appendix 1, Table A4. The majority of evaluations actively involve public end-user. In 83% (166/200) of the evaluations, end-users actively assessed the DHIs using predefined answer options, and in 73% (146/200), the public and patients had the opportunity to individually express their views and emotions regarding the DHIs. The majority of evaluations referenced user-centred approaches (145/200, 72.5%), with UCD being the most commonly mentioned (133/200, 66.5%). Additionally, related approaches such as participatory design, human-centred design, or co-design/creation were mentioned alternatively or in combination with UCD. In 55 evaluations (55/200, 25.5%), none of these user-centred approaches were mentioned.

Discussion

Principal Results

This scoping review provides a holistic overview of the way DHIs for public end-users are evaluated. A total of 160 studies resulting in 200 evaluations were included in this review. It focused on assessing the range of scientific literature concerning various methods, indicators, and types of public involvement. Research in the field of DHI evaluation seems to be in its early stages. Although studies of the last 15 years were eligible, the oldest eligible study is from 2010 and until 2016 the number of studies is in a single-digit range. The majority of eligible publications were from 2020 onwards (101/160, 63.13%), indicating a growing relevance of evaluating DHIs for public end-users. Similar findings are reported in other literature [3,4,6,29].

It is shown that there is a broad range of different DHIs with complex functions so some can not be clearly categorised in established schemes like WHO or NICE.

Furthermore, there is a lack of methodological consistency, and references to existing frameworks such as WHO, NICE, DEHDI, the report of the EXPHO, or the approach of Murray et al., are notably absent. The bare usage of standardised frameworks was also concluded by another review [20]. Instead of these, in more than half of the evaluations standardised usability measurements are used, although most of these measurement instruments are not specially developed for the healthcare context, e.g., SUS. A major usage of standardised measurement methods also resulted from a review of a specific medical field [22]. In other cases, standard measurements are considered insufficient or found to be too complex and in consequence [61], so that self-constructed surveys are often developed. On the one hand, this raises questions about differences regarding the measurement quality, because some self-constructed measurements seem to be not validated. On the other hand, the use of different measurement methods, some validated some not, complicates the comparability of results. Similar results were found in literature/former reviews [29]. In addition, the question arises why current frameworks that were developed by credible institutions or scientists and explicitly focus the healthcare sector are not being referred.

In terms of the evaluation criteria, similarities to the previously described observations can be seen. As with the methodology, no references to existing frameworks like WHO, NICE, DEHDI, report of the EXPHO or the approach of Murray et al., are mentioned. In the few evaluations that refer to existing literature to explain the used terms, mostly ISO 9241-11 is mentioned, although this is also not specially developed for the healthcare context. Regarding the mapping of the extracted large amount of evaluation criteria into the established classification scheme of the WHO, a similar circumstance to the DHI mapping is noticeable. A clear categorisation in established schemes is not possible because of a missing commonly agreed definition and therefore variety of the usage of the evaluated indicators. The lack of a uniform definition and variability in term usage raises concerns about whether this impedes the sharing of evidence among eHealth interventions due to insufficient comparability of results. The same thoughts are found in literature [7,20].

The literature calls for example for UCD processes to increase the involvement of different end-user groups [3]. The results show that the majority of the evaluations referred to user-centred approaches and actively involve public end-users in the evaluation process.

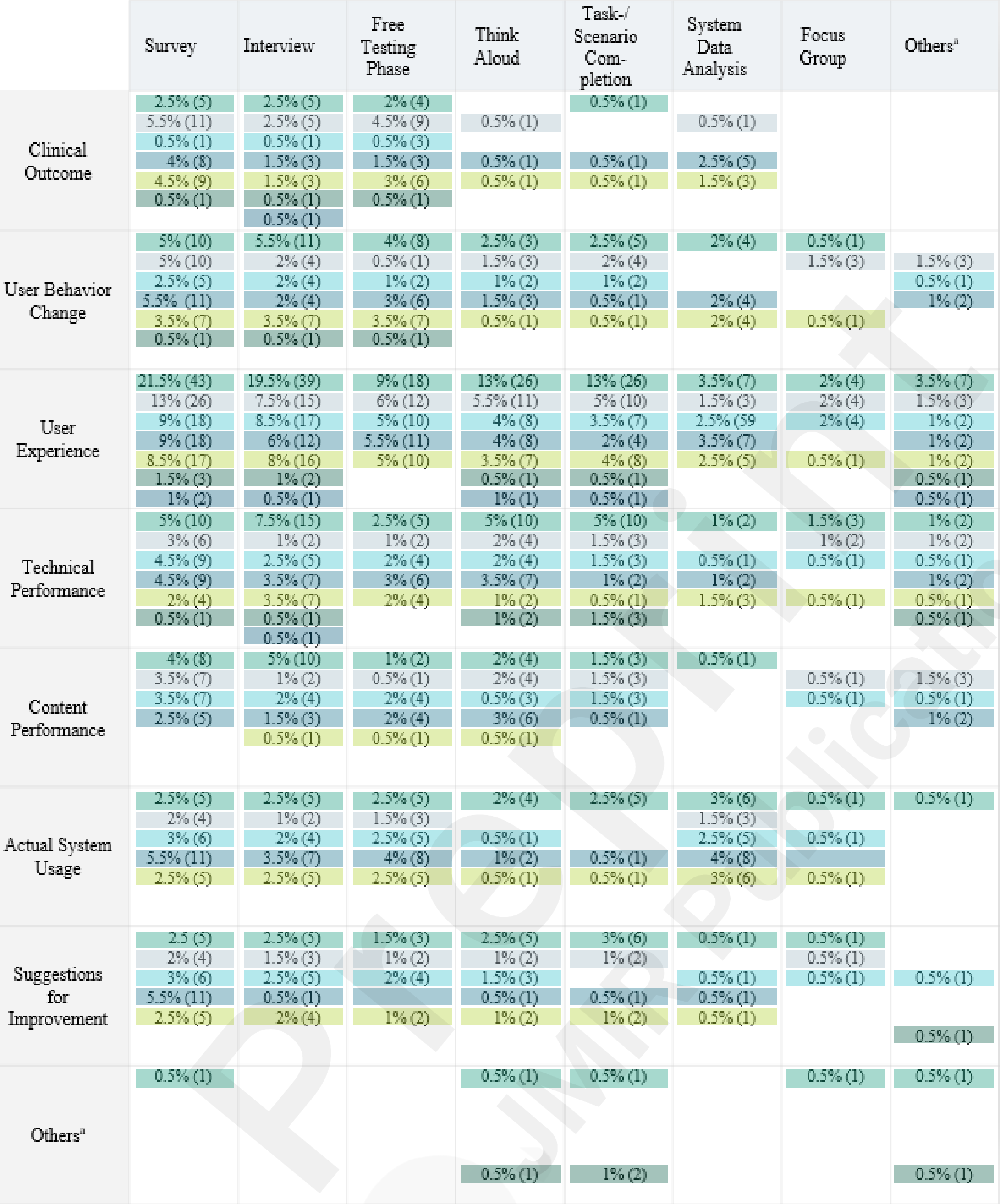
To sum up, the analysis reveals that there is neither a consensus on methods for evaluating DHIs nor a commonly agreed definition or usage of evaluated indicators, resulting in a broad variety of evaluation practices. This aligns with the existing literature [23,62]. Although several frameworks existing the problem of heterogenicity and variability according to the evaluation of DHIs remains and these circumstances seem to hamper gathering reliable evidence through evaluation.

To address these challenges, practical suggestions and implications for further research have been identified and will be described subsequently.

Implications for practice and further research

To consolidate the previously mentioned results on evaluation methods and criteria for each DHI, according to the newly established categorization scheme, Figure 5 presents a results matrix. Considering the observation that established frameworks seem to be not referred, this figure is not intended to function as a framework. Its purpose is to assist developers, evaluators, researchers, and others in this field with the decision-making process by providing an overview of how DHIs have been evaluated by other people. This matrix can address the methodological challenge by serving as initial practical guidance or decision-support for those developing and evaluating DHIs for public end-users, offering a foundation for considering appropriate ways to evaluate their own DHI.

Figure 5- Result matrix- method & criteria per DHI



	Survey	Interview	Free Testing Phase	Think Aloud	Task-/ Scenario Com-pletion	System Data Analysis	Focus Group	Others[a]
Clinical Outcome	2.5% (5)	2.5% (5)	2% (4)		0.5% (1)			
	5.5% (11)	2.5% (5)	4.5% (9)	0.5% (1)		0.5% (1)		
	0.5% (1)	0.5% (1)	0.5% (3)					
	4% (8)	1.5% (3)	1.5% (3)	0.5% (1)	0.5% (1)	2.5% (5)		
	4.5% (9)	1.5% (3)	3% (6)	0.5% (1)	0.5% (1)	1.5% (3)		
	0.5% (1)	0.5% (1)	0.5% (1)					
		0.5% (1)						
User Behavior Change	5% (10)	5.5% (11)	4% (8)	2.5% (3)	2.5% (5)	2% (4)	0.5% (1)	
	5% (10)	2% (4)	0.5% (1)	1.5% (3)	2% (4)		1.5% (3)	1.5% (3)
	2.5% (5)	2% (4)	1% (2)	1% (2)	1% (2)			0.5% (1)
	5.5% (11)	2% (4)	3% (6)	1.5% (3)	0.5% (1)	2% (4)		1% (2)
	3.5% (7)	3.5% (7)	3.5% (7)	0.5% (1)	0.5% (1)	2% (4)	0.5% (1)	
	0.5% (1)	0.5% (1)	0.5% (1)					
User Experience	21.5% (43)	19.5% (39)	9% (18)	13% (26)	13% (26)	3.5% (7)	2% (4)	3.5% (7)
	13% (26)	7.5% (15)	6% (12)	5.5% (11)	5% (10)	1.5% (3)	2% (4)	1.5% (3)
	9% (18)	8.5% (17)	5% (10)	4% (8)	3.5% (7)	2.5% (59	2% (4)	1% (2)
	9% (18)	6% (12)	5.5% (11)	4% (8)	2% (4)	3.5% (7)		1% (2)
	8.5% (17)	8% (16)	5% (10)	3.5% (7)	4% (8)	2.5% (5)	0.5% (1)	1% (2)
	1.5% (3)	1% (2)		0.5% (1)	0.5% (1)			0.5% (1)
	1% (2)	0.5% (1)		1% (1)	0.5% (1)			0.5% (1)
Technical Performance	5% (10)	7.5% (15)	2.5% (5)	5% (10)	5% (10)	1% (2)	1.5% (3)	1% (2)
	3% (6)	1% (2)	1% (2)	2% (4)	1.5% (3)		1% (2)	1% (2)
	4.5% (9)	2.5% (5)	2% (4)	2% (4)	1.5% (3)	0.5% (1)	0.5% (1)	0.5% (1)
	4.5% (9)	3.5% (7)	3% (6)	3.5% (7)	1% (2)	1% (2)		1% (2)
	2% (4)	3.5% (7)	2% (4)	1% (2)	0.5% (1)	1.5% (3)	0.5% (1)	0.5% (1)
	0.5% (1)	0.5% (1)		1% (2)	1.5% (3)			0.5% (1)
		0.5% (1)						
Content Performance	4% (8)	5% (10)	1% (2)	2% (4)	1.5% (3)	0.5% (1)		
	3.5% (7)	1% (2)	0.5% (1)	2% (4)	1.5% (3)		0.5% (1)	1.5% (3)
	3.5% (7)	2% (4)	2% (4)	0.5% (3)	1.5% (3)		0.5% (1)	0.5% (1)
	2.5% (5)	1.5% (3)	2% (4)	3% (6)	0.5% (1)			1% (2)
		0.5% (1)	0.5% (1)	0.5% (1)				
Actual System Usage	2.5% (5)	2.5% (5)	2.5% (5)	2% (4)	2.5% (5)	3% (6)	0.5% (1)	0.5% (1)
	2% (4)	1% (2)	1.5% (3)			1.5% (3)		
	3% (6)	2% (4)	2.5% (5)	0.5% (1)		2.5% (5)	0.5% (1)	
	5.5% (11)	3.5% (7)	4% (8)	1% (2)	0.5% (1)	4% (8)		
	2.5% (5)	2.5% (5)	2.5% (5)	0.5% (1)	0.5% (1)	3% (6)	0.5% (1)	
Suggestions for Improvement	2.5 (5)	2.5% (5)	1.5% (3)	2.5% (5)	3% (6)	0.5% (1)	0.5% (1)	
	2% (4)	1.5% (3)	1% (2)	1% (2)	1% (2)		0.5% (1)	
	3% (6)	2.5% (5)	2% (4)	1.5% (3)		0.5% (1)	0.5% (1)	0.5% (1)
	5.5% (11)	0.5% (1)		0.5% (1)	0.5% (1)	0.5% (1)		
	2.5% (5)	2% (4)	1% (2)	1% (2)	1% (2)	0.5% (1)		
								0.5% (1)
Others[a]	0.5% (1)			0.5% (1)	0.5% (1)		0.5% (1)	0.5% (1)
				0.5% (1)	1% (2)			0.5% (1)

[a] others including (cognitive) walkthrough, observation, informal conversation, design workshops, eye tracking

■ Self-management ■ Digital supportive treatment component ■ Non-tailored information
■ Tailored information ■ Monitoring ■ Communication
■ Data transfer

As long as there is no commonly agreed and broadly applicable approach for evaluating DHIs, certain aspects derived from this review should be considered and described in evaluation studies to enhance the quality and comparability of results, thereby creating evidence. Following the approach of Murray et al., a question-driven approach is chosen for outlining these aspects [1].

What type of DHI is at the centre of the evaluation? What are its main characteristics and (intended) functions?

Describing the characteristics and key components is crucial for comparability and it forms the basis for selecting an appropriate methodology for the study. This description includes determining its primary functions, user interface, the health domain it addresses, the specific medical issue it targets, the population it is designed for and the intended use setting. This aligns with suggestions from existing frameworks [1,5,17]. Moreover, evaluators should reference established classification systems, such as those provided by the WHO or the NICE [5,10]. This approach can help in creating a shared understanding and definitions of the DHI's context, as recommended in the literature [9].

What are the reasons behind choosing the research method? What role do established measurement methods play?
While the description of chosen methods is standard in research, the underlying reasons or decision-making processes for their selection are often not detailed. Therefore, it's essential that evaluators not only carefully select methods suitable for DHI evaluation but also provide clear justifications for these choices [17]. This transparency in decision-making enhances the reliability and validity of the research. Established measurement methods should be given due consideration in the evaluation process and when these are applicable and relevant, they should be integrated into the study. Literature shows different reviews exploring potential frameworks that can be applied to DHIs [18,24]. If established methods are deemed inappropriate or inapplicable for the specific context of the DHI being evaluated, it is important to briefly explain the reasons for their exclusion. This approach not only adheres to high research standards but also helps in understanding the unique aspects of the DHI that necessitate a deviation from conventional methods.

Which criteria are being evaluated? Which established sources or references are considered regarding the evaluated criteria?
Evaluated criteria and the underlying measured indicators should be briefly described by considering established references. Clarity in defining these criteria is essential for creating a standardised and shared understanding of the evaluation's objectives. The selection of criteria should be informed by established references and frameworks in the field of digital health. A possible source to refer to for justification in the context of digitalization and health is the WHO, as it lists and defines criteria such as feasibility, usability, efficacy, and effectiveness [14]. Another valuable reference is the Swiss Evaluation framework by Kowatsch et al. [24], which outlines various criteria specific to DHIs. Referring to established references helps in grounding the evaluation in recognised standards and contributes to the creation of a shared language and understanding in the field, as suggested in the literature [9]. This approach can facilitate comparability across studies and contribute to the broader discourse on DHI efficacy and impact.

In which way are end-users involved in the evaluation process?
In accordance with the NICE framework, it is important to detail how representatives from the intended end-user groups are incorporated into the evaluation [5]. This should include specifying their roles and the extent of their participation, whereby approaches such as PHI and PPI could be referred. Given the potential agile nature of DHI development, these approaches could reach their limits and referring to the principles, strategies, and methods UCD can be highly effective in ensuring meaningful end-user involvement.

The following implication for further research can be derived from the scoping review. During the scoping review, the question arises why current establishes measurements or frameworks, developed by credible institutions or scientists and explicitly focused on the healthcare sector, are not being referenced. This potential research-practice gap could be addressed by further research, for example, by investigating the discrepancies between scientific best practices in user-centred evaluation strategies and the suggestions from established frameworks within the processes of real-world user-centred evaluation. In addition, DHIs are developed not only in academic settings but also in commercial sectors. As stated in limitations, results of evaluation studies are not always published, especially those from the commercial sector. Therefore, it could be interesting to investigate evaluation methods particularly in this field. Same thoughts are found in literature [29]. Regarding the involvement of public end-users, it would be interesting to investigate from a public end-user perspective which methods are suitable for providing close-to-reality feedback. This involves questions such as: How do they like to explore DHIs? Are the measurement methods currently used by evaluators suitable for conveying the benefits to public end-users in a realistic way, thus enabling them to assess the value? From their point of view, which approaches are tangible and can be transferred to the reality of care? Furthermore, it could be interesting to develop and investigate a reporting guideline that ensures an evidence-based minimum set of items for reporting evaluation processes on DHI. Considering the incorporation of potential end-users further research could focus on reporting guideline consisting of standardised components regarding the participation or involvement of end-users in the development or evaluation of DHIs.

Limitations

This scoping review was conducted based on the methodology framework for scoping reviews of Arksey & O'Malley and complies with the PRISMA-ScR guideline. To the best of the authors' knowledge, the results provide a first holistic overview of scientific research on evaluation methods and indicators in the context of DHI for public end-users, because the scope is neither medical-specific nor criteria-specific.

However, the review also has some limitations. One limitation is gathered by the way evaluation studies are published. Generally, evaluations of DHIs are published in one the following two ways: either different evaluations conducted during the development process of a DHI are published in separate papers, or the entire development process is described in a single paper. In the latter case, there are often no detailed descriptions of the evaluation process, as it is not the main focus of the publication. This lack of detail impedes a comprehensive and thorough overview of the evaluation methods and indicators of DHIs throughout their development and actual use. Another limitation is that results of evaluation studies, especially from the commercial sector, are not always published [20,23,29], what also hampers a comprehensive overview. Additionally, in line with the nature of scoping reviews, there is neither a quality assessment of the included studies and evaluations nor a quality assessment of the extracted methods, criteria, and modes of public involvement. These aspects are presented as reported by the authors of the primary studies.

Conclusions

This scoping review provides a comprehensive overview of the current methods used in evaluating DHIs for public end-users. The analysis reveals that there is neither a consensus on methods for evaluating DHIs nor a commonly agreed definition or usage of evaluated indicators, resulting in a broad variety of evaluation practices. Although several frameworks existing the problem of heterogenicity and variability according to the evaluation of DHI remains and these circumstances seem to hamper gathering reliable evidence through evaluation. Recommendations are derived from the findings in order to enhance the quality and comparability of evaluation studies results. It is important to note that the results are not intended to serve as a framework or as best-practice recommendations. Investigating these aspects could form a part of future research endeavours. We demonstrated that the research field is complex, heterogeneous, and broad, and our findings create a first overview and identified research gaps that further could be addressed. In conclusion, a uniform usage of terms, particularly regarding evaluation criteria and DHI classification, within the digital health sector could facilitate the transferability of results among similar evaluation studies. This standardization could significantly contribute to the cohesiveness and effectiveness of research in this evolving field.

Acknowledgements

This research received no external funding.

Conflicts of Interest

none declared.

Abbreviations

DEHDI: Design and evaluation of DHIs
DHI: Digital health intervention
DPHI: Digital public health interventions
EBM: Evidence-based medicine
HTA: health technology assessment
JMIR: Journal of Medical Internet Research
NICE: National Institute for Health and Care Excellence
OSF: Center for Open Science
PHI: Participatory health research
PPI: Public and patient involvement
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomised controlled trial
RQ: Research question
UCD: user-centered design
WHO: World Health Organisation

Multimedia Appendix 1

PRISMA-ScR checklist & classification, search string and categorisation schemes.

Multimedia Appendix 2

Data extraction chart.

References

1. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, Rivera DE, West R, Wyatt JC. Evaluating Digital Health Interventions: Key Questions and Approaches. Am J Prev Med 2016;51(5):843-851. PMID:27745684
2. World Health Organization. WHO Guideline: Recommendations on digital interventions for health system strengthening 2019. https://www.who.int/publications/i/item/9789241550505 [accessed 2023-06-29]. Archived at: 978 92 4 155050 5
3. Fischer F. Digitale Interventionen in Prävention und Gesundheitsförderung: Welche Form der Evidenz haben wir und welche wird benötigt? [Digital interventions in prevention and health promotion: What kind of evidence do we have and what is needed?]. Bundesgesundheitsbl 2020;63(6):674-680. PMID:32355991
4. Gomes M, Murray E, Raftery J. Economic Evaluation of Digital Health Interventions: Methodological Issues and Recommendations for Practice. Pharmacoeconomics 2022;40(4):367-378. doi:10.1007/s40273-022-01130-0
5. National Institute for Health and Care Excellence. Evidence standards framework for digital health technologies; 2022.
6. Zanaboni P, Ngangue P, Mbemba GIC, Schopf TR, Bergmo TS, Gagnon M-P. Methods to Evaluate the Effects of Internet-Based Digital Health Interventions for Citizens: Systematic

Review of Reviews. J Med Internet Res 2018;20(6):e10202. PMID:29880470

7. Kolasa K, Kozinski G. How to Value Digital Health Interventions? A Systematic Literature Review. Int J Environ Res Public Health 2020;17(6). PMID:32209988
8. Kernebeck S, Scheibe M, Sinha M, Fischer F, Knapp A, Timpel P, Harst L, Reininghaus U, Vollmar HC. Digitale Gesundheitsinterventionen entwickeln, evaluieren und implementieren (Teil I) – Diskussionspapier der Arbeitsgruppe Digital Health des Deutschen Netzwerk Versorgungsforschung (DNVF). [Development, Evaluation and Implementation of Digital Health Interventions (Part 1) - Discussion Paper of the Digital Health Working Group of the German Network for Health Services Research (DNVF)]. Gesundheitswesen 2023;85(1):58-64. PMID:36446615
9. Kwasnicka D, Keller J, Perski O, Potthoff S, Hoor GA ten, Ainsworth B, Crutzen R, Dohle S, van Dongen A, Heino M, Henrich JF, Knox L, König LM, Maltinsky W, McCallum C, Nalukwago J, Neter E, Nurmi J, Spitschan M, van Beurden SB, van der Laan LN, Wunsch K, Levink JJJ, Sanderman R. White Paper: Open Digital Health - accelerating transparent and scalable health promotion and treatment. Health Psychol Rev 2022;16(4):475-491. PMID:35240931
10. World Health Organization. 2018. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. https://apps.who.int/iris/handle/10665/260480.
11. Wienert J, Jahnel T, Maaß L. What are Digital Public Health Interventions? First Steps Toward a Definition and an Intervention Classification Framework. J Med Internet Res 2022;24(6):e31921. PMID:35763320
12. Lange O. Health economic evaluation of preventive digital public health interventions using decision-analytic modelling: a systematized review. BMC Health Serv Res 2023;23(1):268. PMID:36932436
13. Maaß L, Pan C-C, Freye M. Mapping Digital Public Health Interventions Among Existing Digital Technologies and Internet-Based Interventions to Maintain and Improve Population Health in Practice: Protocol for a Scoping Review. JMIR Res Protoc 2022;11(3):e33404. PMID:35357321
14. World Health Organization. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment: World Health Organization; 2016. ISBN: 9789241511766.
15. Hochmuth A, Exner A-K, Dockweiler C. Implementierung und partizipative Gestaltung digitaler Gesundheitsinterventionen. [Implementation and participatory design of digital health interventions]. Bundesgesundheitsbl 2020;63(2):145-152. PMID:31938837
16. Hrynyschyn R, Prediger C, Stock C, Helmer SM. Evaluation Methods Applied to Digital Health Interventions: What Is Being Used beyond Randomised Controlled Trials?-A Scoping Review. Int J Environ Res Public Health 2022;19(9). PMID:35564616
17. EXpert Panel on effective ways of investing in Health. Opinion on assessing the impact of digital transformation of health services: Report of the Expert Panel on effective ways of investing in Health (EXPH); 2019.
18. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-A call for innovative evidence generation approaches. npj Digit. Med. 2020;3(1):110. PMID:32904379
19. Huben A von, Howell M, Howard K, Carrello J, Norris S. Health technology assessment for digital technologies that manage chronic disease: a systematic review. International journal of technology assessment in health care 2021;37(1):e66. PMID:34034851
20. Enam A, Torres-Bonilla J, Eriksson H. Evidence-Based Evaluation of eHealth Interventions: Systematic Literature Review. J Med Internet Res 2018;20(11):e10971. PMID:30470678
21. Jahnel T, Schüz B. Partizipative Entwicklung von Digital-Public-Health-Anwendungen:

Spannungsfeld zwischen Nutzer*innenperspektive und Evidenzbasierung. [Participatory development of digital public health: tension between user perspectives and evidence]. Bundesgesundheitsbl 2020;63(2):153-159. PMID:31915866

22. Holl F, Kircher J, Swoboda WJ, Schobel J. Methods Used to Evaluate mHealth Applications for Cardiovascular Disease: A Quasi-Systematic Scoping Review. Int J Environ Res Public Health 2021;18(23). PMID:34886039
23. Kernebeck S, Redaelli M, Kramer U, Vollmar HC. Wie sollen digitale Gesundheitsanwendungen durch die Versorgungsforschung evaluiert werden? 2019. doi:10.3205/19DKVF029
24. Kowatsch T, Harperink S, Cotti A, Trémeaud R, Schneider J, Della Valle M. Evaluierungs-Framework und Kriterienkatalog für digitale Gesundheitsinterventionen in der Schweiz: Gesundheitsförderung Schweiz & Zentrum für digitale Gesundheitsinterventionen der Universität St.Gallen & ETH Zürich; 2021.
25. Karpathakis K, Libow G, Potts HWW, Dixon S, Greaves F, Murray E. An Evaluation Service for Digital Public Health Interventions: User-Centered Design Approach. J Med Internet Res 2021;23(9):e28356. PMID:34494965
26. Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz – DVG): DVG.
27. Hastall MR, Dockweiler C, Mühlhaus J. Achieving End User Acceptance: Building Blocks for an Evidence-Based User-Centered Framework for Health Technology Development and Assessment 2017;10279:13-25. doi:10.1007/978-3-319-58700-4_2
28. Dockweiler C, Razum O. Digitalisierte Gesundheit: neue Herausforderungen für Public Health. [Digital Health: new Challenges for Public Health]. Gesundheitswesen 2016;78(1):5-7. PMID:26695540
29. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. Int J Med Inform 2019;126:95-104. doi:10.1016/j.ijmedinf.2019.03.018
30. International Organization for Standardization. 2023. ISO 9241-210:2019: Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. https://www.iso.org/standard/77520.html.
31. Cornet VP, Toscos T, Bolchini D, Rohani Ghahari R, Ahmed R, Daley C, Mirro MJ, Holden RJ. Untold Stories in User-Centered Design of Mobile Health: Practical Challenges and Strategies Learned From the Design and Evaluation of an App for Older Adults With Heart Failure. JMIR Mhealth Uhealth 2020;8(7):e17703. PMID:32706745
32. Dopp AR, Parisi KE, Munson SA, Lyon AR. Aligning implementation and user-centered design strategies to enhance the impact of health services: results from a concept mapping study. Implement Sci Commun 2020;1(1):17. PMID:32885179
33. International Collaboration for Participatory Health Research. Position Paper 1: What is Participatory Health Research? 2013. http://www.icphr.org/position-papers--discussion-papers/position-paper-no-1.
34. National Institute for Health Research. INVOLVE.
35. National Institute for Health Research (NIHR)- School for Primary Care Research. What is Patient and Public Involvement and Public Engagement? https://www.spcr.nihr.ac.uk/PPI/what-is-patient-and-public-involvement-and-engagement.
36. University of Oxford- Medical Sciences Division. Patient and public involvement, engagement and participation definitions. https://www.medsci.ox.ac.uk/research/patient-and-public-involvement/section-2-what-is-patient-and-public-involvement.
37. National Institute for Health Research. 2014. Going the extra mile: Improving the nation's health and wellbeing through public involvement in research. https://www.nihr.ac.uk/documents/about-us/our-contribution-to-research/how-we-involve-patients-carers-and-the-public/Going-the-Extra-Mile.pdf.

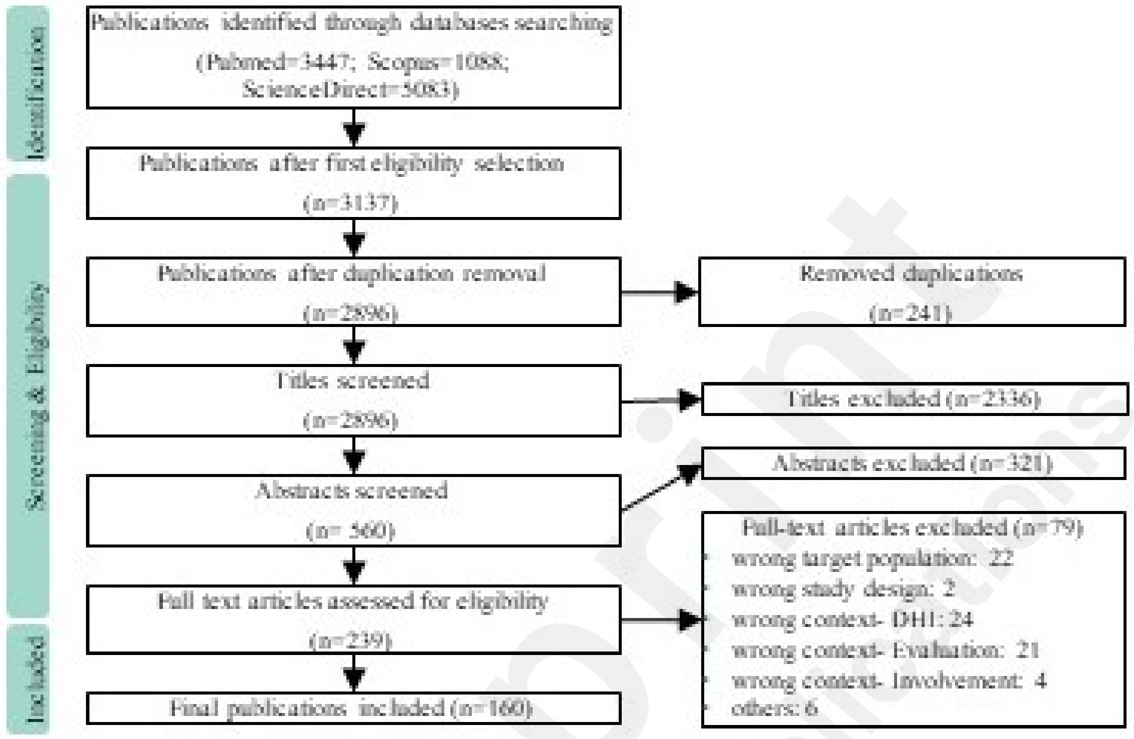
38. Jonsson M, Johansson S, Hussain D, Gulliksen J, Gustavsson C. Development and Evaluation of eHealth Services Regarding Accessibility: Scoping Literature Review. J Med Internet Res 2023;25:e45118. PMID:37590050
39. Scheckel B, Schmidt K, Stock S, Redaèlli M. Patient Portals as Facilitators of Engagement in Patients With Diabetes and Chronic Heart Disease: Scoping Review of Usage and Usability. J Med Internet Res 2023;25:e38447. PMID:37624629
40. Sun Y, Zhou J, Ji M, Pei L, Wang Z. Development and Evaluation of Health Recommender Systems: Systematic Scoping Review and Evidence Mapping. J Med Internet Res 2023;25:e38184. PMID:36656630
41. Elm E von, Schreiber G, Haupt CC. Methodische Anleitung für Scoping Reviews (JBI-Methodologie). [Not Available]. Z Evid Fortbild Qual Gesundhwes 2019;143:1-7. doi:10.1016/j.zefq.2019.05.004
42. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018;18(1):143. PMID:30453902
43. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005;8(1):19-32. doi:10.1080/1364557032000119616
44. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MDJ, Horsley T, Weeks L, Hempel S, Akl EA, Chang C, McGowan J, Stewart L, Hartling L, Aldcroft A, Wilson MG, Garritty C, Lewin S, Godfrey CM, Macdonald MT, Langlois EV, Soares-Weiser K, Moriarty J, Clifford T, Tunçalp Ö, Straus SE. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018;169(7):467-473. PMID:30178033
45. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: Scoping reviews (2020 Version). JBI Manual for Evidence Synthesis 2020. doi:10.46658/JBIMES-20-12
46. Weirauch V. 2023. User-Centered Evaluation of Digital Health Intervention: A Scoping Review. doi:10.17605/OSF.IO/X6GKF
47. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016;23(1):5-11. PMID:26377990
48. Marien S, Legrand D, Ramdoyal R, Nsenga J, Ospina G, Ramon V, Boland B, Spinewine A. A web application to involve patients in the medication reconciliation process: a user-centered usability and usefulness study. J Am Med Inform Assoc 2018;25(11):1488-1500. PMID:30137331
49. Rochat J, Ehrler F, Siebert JN, Ricci A, Garretas Ruiz V, Lovis C. Usability Testing of a Patient-Centered Mobile Health App for Supporting and Guiding the Pediatric Emergency Department Patient Journey: Mixed Methods Study. JMIR Pediatr Parent 2022;5(1):e25540. PMID:35289754
50. Sousa P, Fonseca H, Gaspar P, Gaspar F. Usability of an internet-based platform (Next.Step) for adolescent weight management. Jornal de Pediatria 2015;91(1):68-74. PMID:25245364
51. Mendes-Santos C, Nunes F, Weiderpass E, Santana R, Andersson G. Development and Evaluation of the Usefulness, Usability, and Feasibility of iNNOV Breast Cancer: Mixed Methods Study. JMIR Cancer 2022;8(1):e33550. PMID:35166682
52. Johnson AK, Haider S, Nikolajuk K, Kuhns LM, Ott E, Motley D, Hill B, Hirschhorn L. An mHealth Intervention to Improve Pre-Exposure Prophylaxis Knowledge Among Young Black Women in Family Planning Clinics: Development and Usability Study. JMIR Form Res 2022;6(7):e37738. PMID:35900830
53. Ebenfeld L, Kleine Stegemann S, Lehr D, Ebert DD, Funk B, Riper H, Berking M. A mobile application for panic disorder and agoraphobia: Insights from a multi-methods feasibility study. Internet Interventions 2020;19:100296. PMID:31890640

54. Wiberg A-C, Ghaderi A, Danielsson HB, Safarzadeh K, Parling T, Carlbring P, Jansson M, Welch E. Internet-based cognitive behavior therapy for eating disorders - Development and feasibility evaluation. Internet Interventions 2022;30:100570. PMID:36110307
55. Bendig J, Wolf A-S, Mark T, Frank A, Mathiebe J, Scheibe M, Müller G, Stahr M, Schmitt J, Reichmann H, Loewenbrück KF, Falkenburger BH. Feasibility of a Multimodal Telemedical Intervention for Patients with Parkinson's Disease-A Pilot Study. Journal of Clinical Medicine 2022;11(4):1074. PMID:35207351
56. Ahern DK, Parker D, Eaton C, Rafferty C, Wroblewski J, Goldman R. Patient-facing Technology for Identification of COPD in Primary Care. J Innov Health Inform 2016;23(2):824. PMID:27869583
57. van Doorn M, Nijhuis LA, Monsanto A, van Amelsvoort T, Popma A, Jaspers MWM, Noordzij ML, Öry FG, Alvarez-Jimenez M, Nieman DH. Usability, Feasibility, and Effect of a Biocueing Intervention in Addition to a Moderated Digital Social Therapy-Platform in Young People With Emerging Mental Health Problems: A Mixed-Method Approach. Front. Psychiatry 2022;13:871813. PMID:35693972
58. Shrestha R, Altice FL, DiDomizio E, Sibilio B, Ranjit YS, Copenhaver MM. Feasibility and Acceptability of an mHealth-Based Approach as an HIV Prevention Strategy Among People Who Use Drugs on Pre-Exposure Prophylaxis. PPA 2020;14:107-118. PMID:32021122
59. Peng ML, Wickersham JA, Altice FL, Shrestha R, Azwa I, Zhou X, Halim MAA, Ikhtiaruddin WM, Tee V, Kamarulzaman A, Ni Z. Formative Evaluation of the Acceptance of HIV Prevention Artificial Intelligence Chatbots By Men Who Have Sex With Men in Malaysia: Focus Group Study. JMIR Form Res 2022;6(10):e42055. PMID:36201390
60. Slater H, Stinson JN, Jordan JE, Chua J, Low B, Lalloo C, Pham Q, Cafazzo JA, Briggs AM. Evaluation of Digital Technologies Tailored to Support Young People's Self-Management of Musculoskeletal Pain: Mixed Methods Study. J Med Internet Res 2020;22(6):e18315. PMID:32442143
61. Floch J, Vilarinho T, Zettl A, Ibanez-Sanchez G, Calvo-Lerma J, Stav E, Haro PH, Aalberg AL, Fides-Valero A, Bayo Montón JL. Users' Experiences of a Mobile Health Self-Management Approach for the Treatment of Cystic Fibrosis: Mixed Methods Study. JMIR Mhealth Uhealth 2020;8(7):e15896. PMID:32673237
62. Theodorou P, Tsiligkos K, Meliones A, Filios C. A Training Smartphone Application for the Simulation of Outdoor Blind Pedestrian Navigation: Usability, UX Evaluation, Sentiment Analysis. Sensors 2022;23(1):367. PMID:36616964

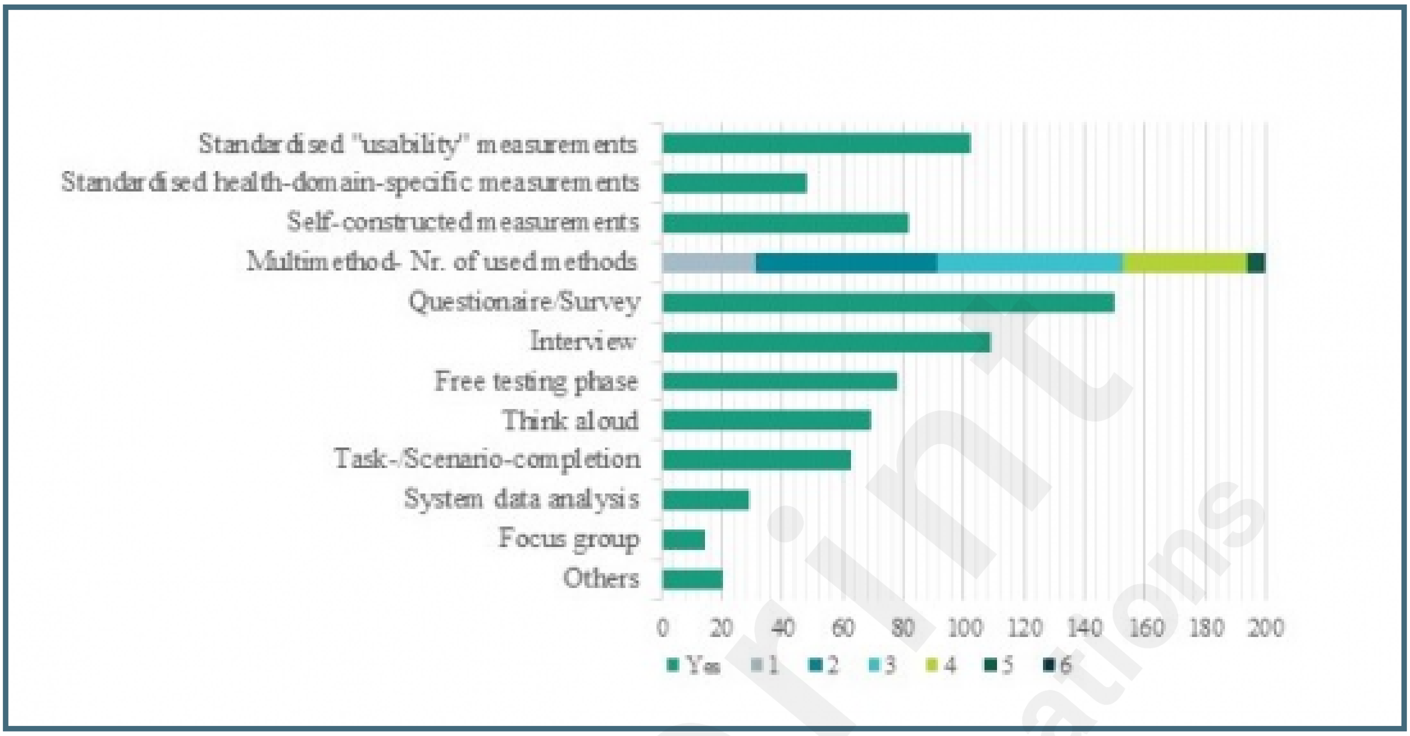
Supplementary Files

Figures

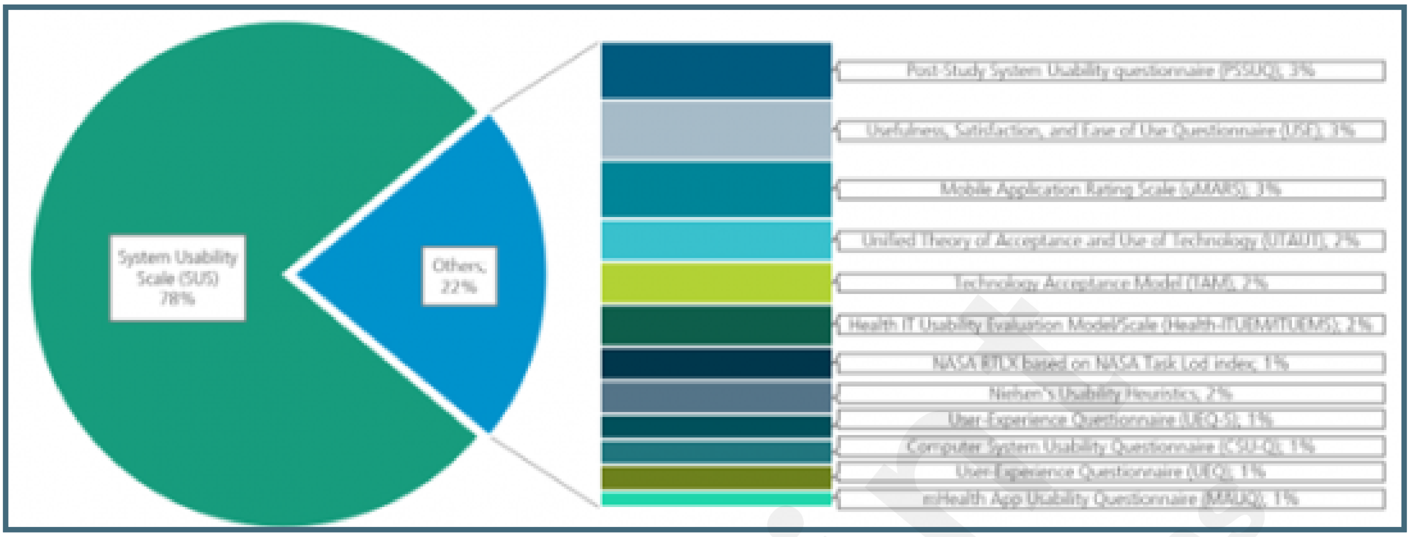
PRISMA workflow chart.



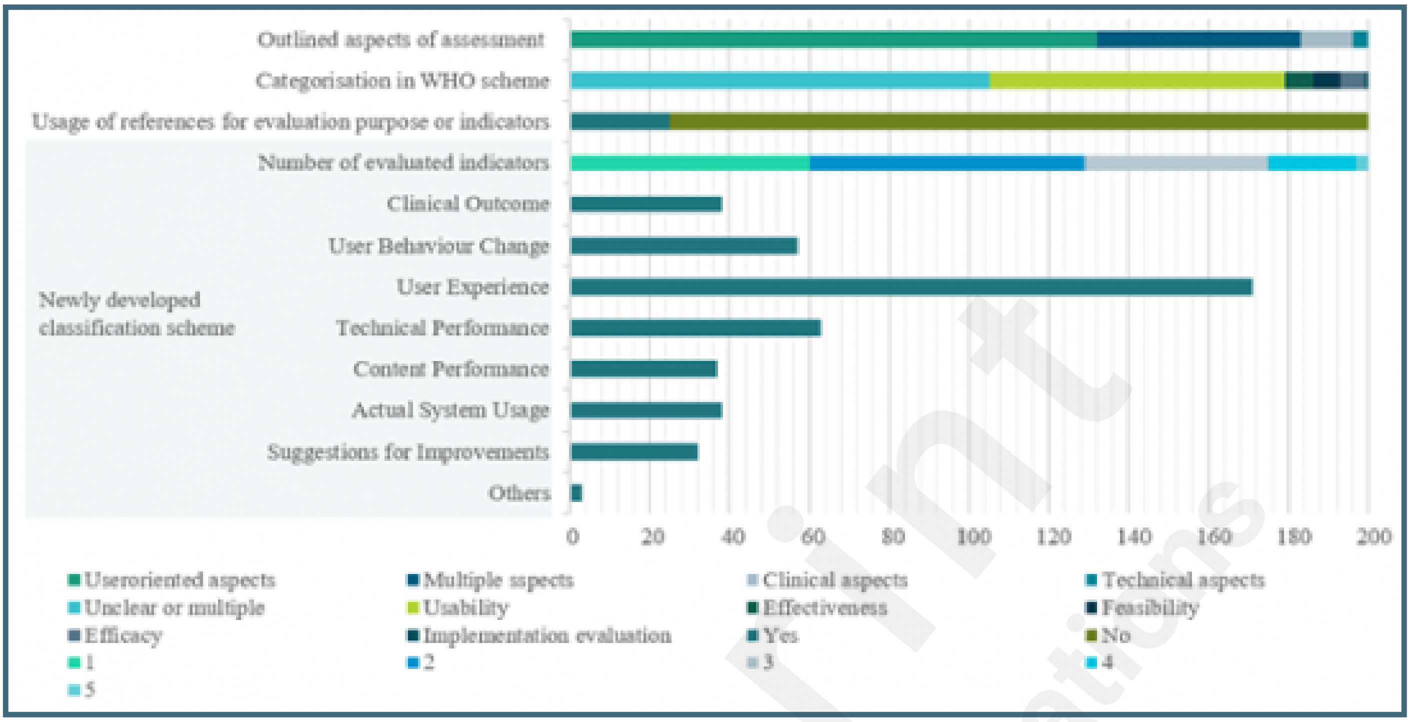
Research methods.



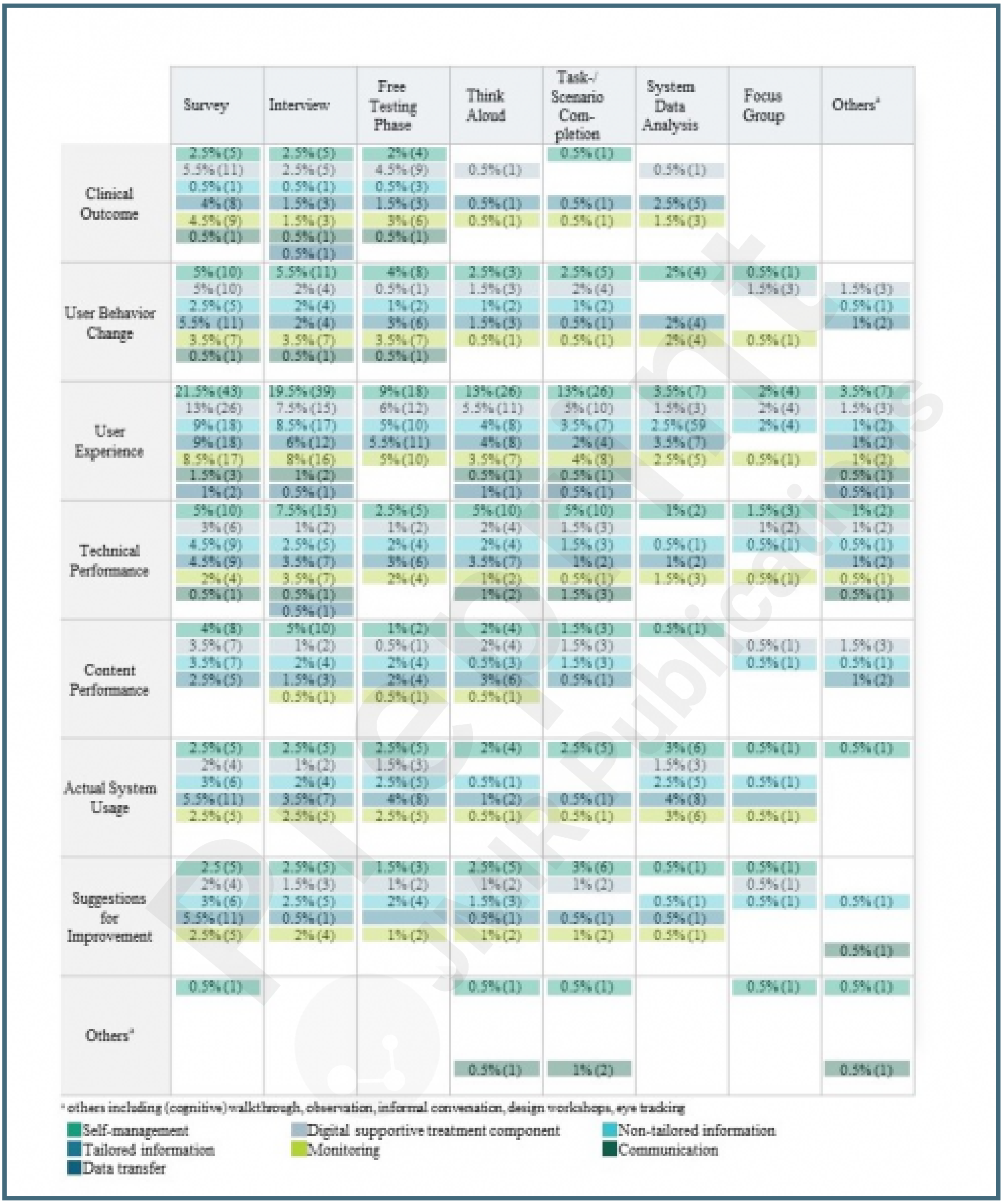
Overview of standardised usability measurements.



Research criteria.



Result matrix- method & criteria per DHI.



	Survey	Interview	Free Testing Phase	Think Aloud	Task-/ Scenario Com-pletion	System Data Analysis	Focus Group	Others[a]
Clinical Outcome	2.5% (5) 5.5% (11) 0.5% (1) 4% (8) 4.5% (9) 0.5% (1)	2.5% (5) 2.5% (5) 0.5% (1) 1.5% (3) 1.5% (3) 0.5% (1) 0.5% (1)	2% (4) 4.5% (9) 0.5% (3) 1.5% (3) 3% (6) 0.5% (1)	0.5% (1) 0.5% (1) 0.5% (1)	0.5% (1) 0.5% (1) 0.5% (1)	0.5% (1) 2.5% (5) 1.5% (3)		
User Behavior Change	5% (10) 5% (10) 2.5% (5) 5.5% (11) 3.5% (7) 0.5% (1)	5.5% (11) 2% (4) 2% (4) 2% (4) 3.5% (7) 0.5% (1)	4% (8) 0.5% (1) 1% (2) 3% (6) 3.5% (7) 0.5% (1)	2.5% (3) 1.5% (3) 1% (2) 1.5% (3) 0.5% (1)	2.5% (5) 2% (4) 1% (2) 0.5% (1) 0.5% (1)	2% (4) 2% (4) 2% (4)	0.5% (1) 1.5% (3) 0.5% (1)	1.5% (3) 0.5% (1) 1% (2)
User Experience	21.5% (43) 13% (26) 9% (18) 9% (18) 8.5% (17) 1.5% (3) 1% (2)	19.5% (39) 7.5% (15) 8.5% (17) 6% (12) 8% (16) 1% (2) 0.5% (1)	9% (18) 6% (12) 5% (10) 5.5% (11) 5% (10)	13% (26) 5.5% (11) 4% (8) 4% (8) 3.5% (7) 0.5% (1) 1% (1)	13% (26) 5% (10) 3.5% (7) 2% (4) 4% (8) 0.5% (1) 0.5% (1)	3.5% (7) 1.5% (3) 2.5% (59 3.5% (7) 2.5% (5)	2% (4) 2% (4) 2% (4) 0.5% (1)	3.5% (7) 1.5% (3) 1% (2) 1% (2) 1% (2) 0.5% (1) 0.5% (1)
Technical Performance	5% (10) 3% (6) 4.5% (9) 4.5% (9) 2% (4) 0.5% (1)	7.5% (15) 1% (2) 2.5% (5) 3.5% (7) 3.5% (7) 0.5% (1) 0.5% (1)	2.5% (5) 1% (2) 2% (4) 3% (6) 2% (4)	5% (10) 2% (4) 2% (4) 3.5% (7) 1% (2) 1% (2)	5% (10) 1.5% (3) 1.5% (3) 1% (2) 0.5% (1) 1.5% (3)	1% (2) 0.5% (1) 1% (2) 1.5% (3)	1.5% (3) 1% (2) 0.5% (1) 0.5% (1)	1% (2) 1% (2) 0.5% (1) 1% (2) 0.5% (1) 0.5% (1)
Content Performance	4% (8) 3.5% (7) 3.5% (7) 2.5% (5)	5% (10) 1% (2) 2% (4) 1.5% (3) 0.5% (1)	1% (2) 0.5% (1) 2% (4) 2% (4) 0.5% (1)	2% (4) 2% (4) 0.5% (3) 3% (6) 0.5% (1)	1.5% (3) 1.5% (3) 1.5% (3) 0.5% (1)	0.5% (1)	0.5% (1) 0.5% (1)	1.5% (3) 0.5% (1) 1% (2)
Actual System Usage	2.5% (5) 2% (4) 3% (6) 5.5% (11) 2.5% (5)	2.5% (5) 1% (2) 2% (4) 3.5% (7) 2.5% (5)	2.5% (5) 1.5% (3) 2.5% (5) 4% (8) 2.5% (5)	2% (4) 0.5% (1) 1% (2) 0.5% (1)	2.5% (5) 0.5% (1) 0.5% (1)	3% (6) 1.5% (3) 2.5% (5) 4% (8) 3% (6)	0.5% (1) 0.5% (1) 0.5% (1)	0.5% (1)
Suggestions for Improvement	2.5 (5) 2% (4) 3% (6) 5.5% (11) 2.5% (5)	2.5% (5) 1.5% (3) 2.5% (5) 0.5% (1) 2% (4)	1.5% (3) 1% (2) 2% (4) 1% (2)	2.5% (5) 1% (2) 1.5% (3) 0.5% (1) 1% (2)	3% (6) 1% (2) 0.5% (1) 1% (2)	0.5% (1) 0.5% (1) 0.5% (1) 0.5% (1)	0.5% (1) 0.5% (1) 0.5% (1)	0.5% (1) 0.5% (1)
Others[a]	0.5% (1)			0.5% (1) 0.5% (1)	0.5% (1) 1% (2)		0.5% (1)	0.5% (1) 0.5% (1)

[a] others including (cognitive) walkthrough, observation, informal conversation, design workshops, eye tracking

■ Self-management
■ Tailored information
■ Data transfer
■ Digital supportive treatment component
■ Monitoring
■ Non-tailored information
■ Communication

Multimedia Appendixes

PRISMA-ScR checklist & classification, search string and categorisation schemes.
URL: http://asset.jmir.pub/assets/a865d55be6d4094be210bf66c922996c.docx

Data extraction chart.
URL: http://asset.jmir.pub/assets/e1c6e3f743deabab2c380831cff4d2b7.xlsx