Evaluating the Digital Patient Experience: Timing, Indicators, and Approaches Based on an Umbrella Systematic Review

Tingting Wang, Guido Giunti, Richard Goossens, Marijke Melles

Submitted to: Journal of Medical Internet Research
on: February 06, 2023

Table of Contents

Evaluating the Digital Patient Experience: Timing, Indicators, and Approaches Based on an Umbrella Systematic Review

Tingting Wang[1] MA; Guido Giunti[2] PhD; Richard Goossens[3] PhD; Marijke Melles[3] PhD

[1]Delft University of Technology DELFT NL
[2]University of Oulu Oulu FI
[3]Delft University of Technology Delft NL

Corresponding Author:
Tingting Wang MA
Delft University of Technology
Michiel de Ruyterweg 128
DELFT
NL

Abstract

Background: Despite the growing number of digital health (DH) applications, timely, cost-effective, and robust evaluations have not kept pace. Patient experience (PEx) was reported as one of the challenges facing the health system by the World Health Organization (WHO) in its 2018 classification of digital health initiatives report. To generate evidence of DH and promote the appropriate integration and use of technologies, standard evaluation of PEx in DH is needed.

Objective: In this study, we systematically reviewed the literature on PEx evaluation timing considerations, evaluation indicators, and evaluation approaches in DH and generated an evaluation guide for further measurement of PEx in DH.

Methods: We performed an umbrella systematic review following PRISMA, searching Scopus, PubMed, and Web of Science databases. Two rounds of small random sampling (20%) were independently reviewed by two reviewers who evaluated the eligibility of the articles against the selection criteria. Two-round interrater reliability was assessed using the Fleiss-Cohen coefficient (k1=0.88 and k2=0.80). Thematic analysis was then applied to analyze the extracted data based on a set of a priori categories.

Results: The search yielded 173 records, of which 45 (26%) were eligible for analysis. Our review highlights five typical evaluation objectives which serve five stakeholder groups separately. We identified three evaluation timing considerations and classified them into three categories: intervention maturity stages, timing of the evaluation, and timing of data collection. Information on evaluation indicators of the digital PEx was identified and summarized into 3 categories (intervention outputs, patient outcomes, and healthcare system impact), 9 themes, and 22 subthemes. A set of evaluation theories, common study designs, data collection methods and instruments, and data analysis approaches were captured which can be used or adapted to evaluate the digital PEx.

Conclusions: Our findings enabled us to generate an evaluation guide to help digital health interventions (DHI) researchers, designers, developers, and program evaluators to evaluate the digital PEx. Finally, we propose six directions for encouraging further digital PEx evaluation research and practice to address the poor PEx challenge.

(JMIR Preprints 06/02/2023:46308)
DOI: https://doi.org/10.2196/preprints.46308

✓ **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

Tingting Wang[1], MA; Guido Giunti[1,2,3], PhD; Richard Goossens[1], PhD; Marijke Melles[1], PhD
[1]*Delft University of Technology, Delft, the Netherlands*
[2]*University of Oulu, Oulu, Finland*
[3]*Trinity College Dublin, Dublin, Ireland*

Corresponding Author:

Tingting Wang, MA
Delft University of Technology
Gebouw 32
Landbergstraat 15
2628 CE Delft
The Netherlands
Phone: +31 623018218
Email: t.wang-8@tudelft.nl

Evaluating the Digital Patient Experience: Timing, Indicators, and Approaches Based on an Umbrella Systematic Review

Abstract

Background: Despite the growing number of digital health (DH) applications, timely, cost-effective, and robust evaluations have not kept pace. Patient experience (PEx) was reported as one of the challenges facing the health system by the World Health Organization (WHO) in its 2018 classification of digital health initiatives report. To generate evidence of DH and promote the appropriate integration and use of technologies, standard evaluation of PEx in DH is needed.

Objectives: The objectives of this study are to systematically identify (1) evaluation timing considerations (i.e., *when to measure*), (2) evaluation indicators (i.e., *what to measure),* and (3) evaluation approaches (i.e., *how to measure)* regarding digital PEx. The overall study aim is to generate an evaluation guide for further improving digital PEx evaluation research and practice.

Methods: This is a parallel study that consists of two phases to our previous study; in phase 1, we followed the same study search and selection process in the current study but focused on a different data extraction and analysis process. Literature reviews related to patient experience in digital health were systematically searched from Scopus, PubMed, and Web of Science. After removing duplicates, we collected 173 records. Two independent raters conducted two rounds of paper screening, including 1) title and abstract screening and 2) full-text screening, and assessed the inter-rater reliability for 20% (n1=23 and n2=12) random samples using the Fleiss-Cohen coefficient (k1=0.88 and k2=0.80). When reaching inter-rater reliability (k>0.60), the first author conducted the rest of the screening process, leaving any uncertainties for group discussions. 45 (26%) articles met the criteria and were considered eligible for further analysis. Subsequently, we applied thematic analysis to the collected data related to research objectives, categorizing it based on a predetermined set of themes on evaluation time considerations, evaluation indicators, and evaluation approaches. In phase 2, to check if there were any meaningful novel insights that would change our conclusions, we performed an updated literature search where we collected 294 newly published reviews, of which 102 were identified as eligible articles. We considered them to have no significant changes to our original results on the research objectives. Therefore, they were not integrated into the synthesis of the current review and were used as supplementary materials in the appendices.

Results: Our review highlights five typical evaluation objectives which serve five stakeholder groups separately. We identified a set of key evaluation timing considerations and classified them into 3 categories: intervention maturity stages, timing of the evaluation, and timing of data collection. Information on evaluation indicators of the digital PEx was identified and summarized into 3 categories (intervention outputs, patient outcomes, and healthcare system impact), 9 themes, and 22 subthemes. A set of evaluation theories, common study designs, data collection methods and instruments, and data analysis approaches were captured which can be used or adapted to evaluate the digital PEx.

Conclusions: Our findings enabled us to generate an evaluation guide to help digital health interventions (DHI) researchers, designers, developers, and program evaluators to evaluate the digital PEx. Finally, we propose six directions for encouraging further digital PEx evaluation research and practice to address the poor PEx challenge.

Keywords: digital health**;** eHealth; telemedicine; mHealth; patient experience; user experience; evaluation timing; evaluation indicators; evaluation approaches; user-centered design; patient-centered care; human-computer interaction

Introduction

Emerging digital technologies promise to shape the future healthcare industry [1, 2]. According to our previous review study [3], most researchers had a positive impression of digital health interventions (DHIs). The number of DHIs proliferates [4-6], which is affecting the way patients receive their healthcare services compared to face-to-face healthcare services and ultimately influencing the patient journey and overall patient experience (PEx) [7, 8]. PEx is a key intent of patient-centered care [9] and a core measure of care quality in digital health [10, 11]. Digital technologies have the potential to enhance or provide comparable PEx compared to some face-to-face healthcare services [8, 12-14]. However, the uptake of digital technologies in healthcare is not as rapid as it has been in many other industries [15], and their potential in healthcare remains unfulfilled [16]. According to the World Health Organization's (WHO) report on the classification of DHIs, the health system is not responding adequately to the need for improved PEx [17].

Despite the growing number of DHIs, timely, cost-effective, and robust evaluations have not matched the growth in numbers [7, 18, 19]. Patient experiences in the wide range of DHIs are mixed [20, 21]. Few published DHIs have resulted in high download numbers and active users [22]; most are released with minimal or no evaluation and require patients to assess the quality for themselves and take responsibility for any consequences [23]. Low-quality digital health may disrupt user experience (UX) [24], resulting in low acceptance, and some may even be harmful [25]. In addition, a DHI may be popular with patients but not valued by clinicians [26]. To generate evidence and promote the appropriate integration and use of digital technologies in healthcare, an overview of how to evaluate PEx or UX in varied DHIs is needed [3, 27].

Evaluating the Digital Patient Experience

In the current study, we use the definition of the digital patient experience from our previous review [3]: “the sum of all interactions affected by a patient’s behavioral determinants, framed by digital technologies, and shaped by organizational culture, that influence patient perceptions across the continuum of care channeling digital health”. This incorporates influencing factors of the digital PEx [3] and the existing definitions of DHIs [28, 29], PEx [30], and UX [31]. Compared to the general PEx and UX, it highlights patient perceptions that are affected by technical, behavioral, and organizational determinants when interacting with a DHI. DHI has become an umbrella term that often encompasses broad concepts and technologies [32], such as DH applications, ecosystems, and platforms [28]. In the current paper, we follow WHO’s definition of DHIs [29]: the use of digital, mobile, and wireless technologies to support the achievement of health objectives. It refers to the use of information and communication technologies (ICT) for healthcare, encompassing both mHealth and eHealth [29, 33]. Compared to evaluating DHIs, PEx, and UX, little is known about evaluating the digital PEx. However, combining the definition of the digital PEx with the extensively explored measurement of PEx, UX, and DHIs can lead to an improved understanding of and enable the development of evaluation approaches for measuring the digital PEx. Therefore, the evaluations of PEx, UX, and DHIs will be used as a starting point in the current paper to clarify when to measure, what to measure, and how to measure the digital PEx.

When to measure

First, the timing of measuring and evaluating digital PEx is an important consideration and must be appropriate for the contextual situation, to make both practical and purposeful sense [34, 35]. According to the European Union [36] and the Department of Health of The King's Fund [37], an evaluation can be scheduled during the design phase, or during or after the implementation phase. Similarly, the WHO [29] introduced three DHI's evaluation stages: efficacy, effectiveness, and implementation. These concepts sequentially refer to where the intervention is under highly controlled conditions, is implemented in a real-world context, and occurs after efficacy and effectiveness have been established. Furthermore, an evaluation can be performed before, during, or after the evaluated intervention in both research and non-research settings [36]. However, decision-making on 'when' to collect PEx data can be more complicated. As argued in earlier studies [35, 37], immediate feedback has the benefit of gaining real-time insights, but patients may be too unwell, stressed or distracted to provide detailed opinions. On the other hand, when the feedback is related to medical outcomes or quality of life, it often requires a lengthy period after the intervention to observe any changes. But responses gathered long after a care episode may be inferior due to recall bias.

What to measure

Second, there is need for a decision on 'what' is required to measure to assess digital PEx. The frequently mentioned UX evaluation concepts, such as usability, functionality, and reliability, from studies [38-40] investigating UX can be applied to evaluate the intervention outputs to anticipate the digital PEx at a service level. Besides, according to the existing constructs and frameworks of understanding or evaluating the PEx [41-45], such as emotional support, relieving fear and anxiety, patients as active participants in care, and continuity of care and relationships, they can be adjusted to evaluate the digital PEx by understanding patient outcomes at an individual level. In addition, the National Quality Forum (NQF) [11] proposed a set of measurable concepts to be used to evaluate PEx in telehealth, for example, patients' increased confidence in, understanding of, and compliance with their care plan; reduction in diagnostic errors and avoidance of adverse outcomes; and decrease in waiting times and eliminated travel. Some of them can be used to understand digital PX at an organizational level by assessing the healthcare system's impact.

How to measure

The third consideration needed is to choose evaluation approaches appropriate for evaluating the digital PEx [35] starting from widely used theories, study designs, methods, and tools for evaluating DHIs and the related PEx or UX. There is rapidly evolving guidance for guiding DH innovators [18], such as the NICE Evidence Standards Framework for Digital Health Technologies [46]. The strength of the evidence in the DHIs' evaluation often depends upon the study design [18]. However, the high bar for evidence in healthcare usually requires longer time for evidence generation, such as prospective RCTs and observational studies, which often conflicts with the fast-innovation reality of the technology industry [16, 18]. Additionally, many traditional approaches, such as qualitative and quantitative methods, can be used to collect experience-related data to evaluate the DHIs [18, 29]. Qualitive methods like focus group, interviews, and observations are often used to obtain an in-depth understanding of PEx [37], in the early intervention development stages [29]. Surveys using structured questionnaires, such as patient satisfaction ratings [37, 47], patient-reported experience measures (PREMs) [35, 48], and patient-reported outcome measures (PROMs) [35, 37, 48] are often used to examine patterns and trends from a large sample. Philip Hodgson believed that strong evidence results from UX data that are valid and reliable, such as formative and summative usability tests, and stated that behavioral data are strong, but opinion data are weak [49].

Objectives

The objectives of this study are to systematically identify (1) evaluation timing considerations (i.e., *when to measure*), (2) evaluation indicators (i.e., *what to measure)*, and (3) evaluation approaches (i.e., *how to measure)* regarding digital PEx. The overall study aim is to generate an evaluation guide for further improving digital PEx evaluation research and practice.

Methods

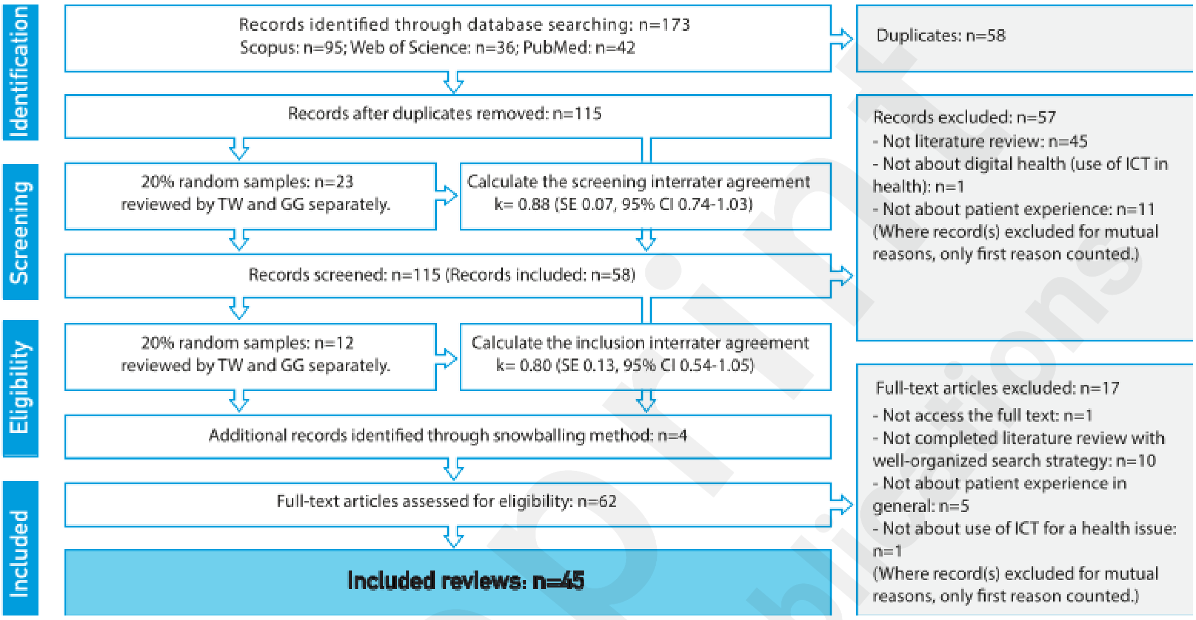
This part consists of two phases. In phase 1, we followed the same study search and selection process as our previous research [3] but focused on a different data extraction and analysis process to achieve our objectives in the current study. In the previous study, we identified the influencing factors and design considerations of the digital PEx, provided a definition, constructed a design and evaluation framework, and generated nine design guidelines to help digital health designers and developers improve the digital PEx. To highlight the connections between ‘design’ and ‘evaluation’ works in the development of digital health and provide readers with a clear roadmap, we included some evaluation-related information in the previous paper as well. However, it was limited and described at a very abstract level. In the current study, detailed information on the evaluation was provided, including evaluation timing considerations, evaluation indicators, and evaluation approaches, and we aimed to generate an evaluation guide for improving the measurement of digital PEx. Given that this is an evolving area, after we finished phase 1, we conducted an updated literature search as a subsequent investigation to decide whether an update of a review was needed in the current study.

Phase 1: The Original Review

Study Search and Selection

Following the PRISMA guideline [50], we conducted an umbrella systematic review [51] on literature reviews that related to PEx and UX in DH. The term DH was first introduced in 2000 by Seth Frank [52]. Therefore, Scopus, PubMed, and Web of Science were used as databases for searching related articles that were published between January 1, 2000, and December 16, 2020. Furthermore, Google Scholar was used to search for additional studies that were identified during the review process through the snowballing method. The computer search resulted in 173 articles, with 58 (33.5%) being duplicates. After removing the duplicates, the titles and abstracts of a small random sampling (23/115, 20%) were reviewed by two independent raters to assess the interrater reliability by using the Fleiss-Cohen coefficient, which resulted in k1=0.88 (SE 0.07, 95% CI 0.74-1.03). Followed by a group discussion to reach agreement on the selection criteria. Subsequently, the remaining titles and abstracts (92/115, 80%) were reviewed by the first author individually. After the titles and abstracts screening, half of the articles (58/115, 50.4%) remained for the full-text review. Meanwhile, four additional articles were identified through snowballing and included in the full-text screening as well. Another small random sampling (12/62, 20%) was reviewed by the two raters to screen the full texts. After achieving interrater reliability: k2=0.80 (SE 0.13, 95% CI 0.54-1.05) and reaching a consensus on the inclusion criteria through another group discussion, the first author reviewed the full texts of the remaining papers (50/62, 80%). Google Sheets was used for performing the screening process and assessments. Finally, as Figure 1 [3] shows, 45 articles were included for data extraction. A detailed search strategy, selection criteria, and screening process can be found in our previous published parallel study [3]. Multimedia Appendix 1 presents the included and excluded articles.

Figure 1. Study flow diagram. ICT: information and communication technology [3].



Data Extraction and Thematic Analysis

We used ATLAS.ti (Scientific Software Development GmbH; version 9.0.7; 1857) for data extraction. Data were extracted for the three predefined objectives: (1) evaluation timing considerations, (2) evaluation indicators, and (3) evaluation approaches of the digital PEx. In addition, we also collected data related to evaluation objectives among the included studies. Data analysis followed Braun and Clarke’s six-phase thematic analysis method [53, 54]: familiarization, coding, generating themes, reviewing themes, defining and naming themes, writing up. First, we became familiar with the 45 included articles. Second, after a thorough review, the first author started iteratively coding the data related to the predefined objectives based on existing frameworks, including the PRISM framework [55], M&E guide [29], measures of PEx in hospitals [37], and an overview of research methodology [56]. This resulted in 25 initial codes. After no more new codes were identified, the first author proposed a coding scheme to summarize recurring points throughout the data. Then the other authors reviewed and discussed the coding scheme until they reached an agreement. Third, the first author followed the coding scheme to code the data more precisely and completely and searched for themes among the generated codes, Fourth, again, the author team reviewed and discussed these codes and themes to address any uncertainties. Fifth, the definitions and names of the generated themes were adjusted through team discussions. Last, the analytical themes related to evaluation timing, indicators, and approaches were produced and reported. Both deductive and inductive approaches [54] were used to identify and generate themes. Four researchers were involved in the review process.

We first highlighted evaluation timing considerations in terms of *intervention maturity stages*, *the timing of evaluation, and the timing of data collection*, which adopted from the description of the

WHO and European Union (see Table 1) [36, 37].

Table 1. Initial Codes of Evaluation Timing considerations of the Digital Patient Experience

Categories and initial codes	Description
Intervention maturity stages [29, 36, 37]	
Efficacy	Assess whether the DHI achieves the intended results in research or controlled setting.
Effectiveness	Assess whether the DHI achieves the intended results in non-research or uncontrolled setting.
Implementation	Assess the uptake, institutionalization and sustainability of evidence based DHIs in a given context, including policies and practices.
Timing of the evaluation [36]	
Before intervention	A pre-test is performed before individuals adopt or implement the intervention. It assesses individuals' initial status and their anticipated perception of the intervention.
During intervention	An evaluation performed during the use of intervention is to monitor individuals' real-time feedback and reactions.
After intervention	The post-test is performed right after or a long time after the completion of the interventions by individuals. It assesses individuals' changes regarding of using the intervention.
Timing of data collection [35, 37]	
Immediate evaluation	Aims to collect "real-time" data on patients' experiences during or immediately after their treatment.
Delayed evaluation	Aims to obtain more substantial responses after the intervention's completion over a long period of time.
Momentary evaluation	Aims to collect transient information from individuals at a specific moment.
Continuous evaluation	Aims to gather feedback from individuals at different points along the care pathway.

We then determined the evaluation indicators, classifying them into three categories (Table 2). *Intervention outputs* are the direct products or deliverables of process activities and refer to the different stages of evaluation that correspond to the various stages of maturity of the DHI. *Patient outcomes* describe the intermediate changes of patients, including patients' emotions, perceptions, capabilities, behaviors, and health conditions as determined by DHIs in terms of influencing factors and interaction processes. *Healthcare system impact* is the medium- to long-term, large-scale financial (intended and unintended) effects produced by a DHI.

Table 2. Initial Codes of Evaluation Indicators of the Digital Patient Experience

Categories and initial codes	Description
Intervention outputs [29, 38-40, 57]	
Functionality	Assess whether the DHI works as intended. It refers to the ability of the DH system to support the desired intervention.
Usability	Assess whether the DHI is used as intended. It refers to the degree to which the intervention is understandable and easy to use.
Quality of care	Assess whether the DHI delivers effective, safe, people-centered, timely, accessible, equitable, integrated, efficient care services. It refers to the degree to which health services for individuals and populations increase the likelihood of desired health outcomes.
Patient outcomes [11, 41-45]	
Emotional outcomes	Assess whether patients' feelings and well-being change positively or negatively because of the use or anticipated use of DHIs. It refers to what

	the patients feels.
Perceptual outcomes	Assess whether the informed state of mind that patients achieve as intended before, during or after using the DHIs. It refers to what the patient thinks and believes.
Capability outcomes	Assess whether patients' health literacy, communication skills, or computer confidence in managing diseases, communicating with health care providers (HCPs), or operating digital devices increased as expected. It refers what the patient knows and acquires.
Behavior outcomes	Assess whether patients entail activities that patients act out to face the disease and the treatments due to DHIs. It refers to what the patient acts and does.
Clinical outcomes	Assess whether patients' health improvements meet the intentions of the DHIs. It refers to what medical condition the patient is in and aims to maintain.
Healthcare System Impact [29]	
Economic outcomes	Assess whether the DHIs are cost-effective, whether the organization and DH users can afford the DHI system, and whether there is a probable return on investment. It refers to the utilization of healthcare resources.

Last, we concluded evaluation approaches in terms of study designs, data collection methods and instruments, and data analysis approaches (Table 3). According to the WHO [29], study designs are intended to assist in decision-making on evidence generation and clarify the scope of evaluation activities. Data collection and analysis are designed through an iterative process that involves strategies for collecting and analyzing data, and a series of specifically designed tools [36].

Table 3. Initial Codes of Evaluation Approaches of the Digital Patient Experience

Categories and initial codes	Description
Study designs [29]	
Descriptive study	Aims to define the "who, what, when, and where" of observed phenomena and include qualitative research concerning both individuals and populations.
Analytical study	Aims to quantify the relationship between the intervention and the outcomes of interest, usually with the specific aim of demonstrating a causative link between the two, including experimental and observational studies.
Data collection methods and instruments [58]	
Qualitative methods	Qualitative research is expressed in words. It is used to understand concepts, thoughts, or experiences. Common qualitative methods include interviews with open-ended questions, observations described in words, and literature reviews that explore concepts and theories.
Quantitative methods	Quantitative research is expressed in numbers and graphs. It is used to test or confirm theories and assumptions. Common quantitative methods include experiments, observations recorded as numbers, and surveys with closed-ended questions.
Data analysis approaches New table	

Data collection methods	Studies, n (%)	References
Questionnaires	33 (73.3%)	[59, 61, 62, 64, 65, 67, 68, 70-75, 77, 78, 81, 82, 85-93, 95-99, 101]
Surveys	32 (71.1%)	[59, 61, 64, 65, 67-75, 77-80, 82, 83, 85-89, 92, 93, 95-97, 99-101]
Interviews	31 (68.9%)	[59, 61, 62, 65-68, 71-75, 77-79, 81-84, 86-93, 97, 98, 101, 105]
Focus groups	19	[61, 62, 64-66, 69, 72-74, 82, 84, 86, 87, 89,

	(42.2%)	91-93, 101, 105]
Observations	17 (37.8%)	[61, 66, 72, 73, 75, 77, 78, 82, 84, 86, 90-94, 97, 105]
Log data	13 (28.9%)	[61, 67, 75, 77-80, 87, 89, 90, 96, 101, 105]
Open-ended questions	10 (22.2%)	[59, 62, 65, 68, 71, 73, 81, 83, 86, 92]
Likert scales	10 (22.2%)	[59, 64, 71, 73, 77, 88, 90, 95, 97, 99]
Usability testing	8 (17.8%)	[59, 63, 66, 70, 73, 87-89]
Diaries	6 (13.3%)	[59, 61, 74, 86, 96, 105]
Contextual inquiry	5 (11.1%)	[59, 62, 75, 86, 93]
Needs assessment	5 (11.1%)	[59, 83, 88, 89, 93]
Performance tests	5 (11.1%)	[66, 67, 71, 78, 90]
Field notes	4 (8.9%)	[62, 75, 91, 105]
Workshops	4 (8.9%)	[73, 74, 88, 89]
Forms	3 (6.7%)	[59, 78, 88]
Think-aloud	3 (6.7%)	[59, 74, 75]
Benchmark testing	2 (4.4%)	[67, 93]
Human impact assessment methodologies	1 (2.2%)	[101]
Personas	1 (2.2%)	[93]

Qualitative analysis	Qualitative data consists of text, images, or videos instead of numbers. Content analysis, thematic analysis, and discourse analysis are common approaches used to analyze this type of data.
Quantitative analysis	Quantitative data is based on numbers. Simple math or more advanced statistical analysis is used to discover commonalities or patterns in the data.

Phase 2: The Updated Scoping Search

The decision to undertake an update of a review requires many considerations. Review authors should consider both whether an update for a review is necessary and when it will be more appropriate [59]. In the light of the “decision framework to assess systematic reviews for updating, with standard terms to report such decisions” [60], we consider that research on patient experience in digital health remains important and evolves rapidly. In case we missed some newly published articles that would bring significant changes to our initial findings, we conducted a rapid scoping search for articles published after our last search. We re-ran the search strategy as specified before with the addition of date (from December 16, 2020, to August 18, 2023) limits set to the period following the most recent search. After removing duplicates (n=73), we collected 294 articles in total. Following the same screening process and selection criteria, we finally identified 102 new eligible articles. The excluded articles were either not a literature review with systematic search (n=74), not about digital health (n=87), not about patient experience (n=26), our own parallel publications (n=2), or not accessible in full text (n=3). The eligible and ineligible articles at this phase can be found in Multimedia Appendix 2. We found that new studies observed results consistent with the existing data. For example, these articles either aim to investigate what factors influence the feasibility, efficacy, effectiveness, design, and implementation of digital health, examine how patients expect, perceive, and experience the digital health interventions, or intend to compare the digital health interventions with conventional face-to-face healthcare services. The research objectives of these new eligible articles can be found in Multimedia Appendix 3. We considered that their findings were unlikely to meaningfully impact our findings on when to measure, what to measure, and how to measure the digital PEx. As suggested by Cumpston M. and Flemyng E. [59]: review authors should decide whether and when to update the review based on their expertise and individual assessment of the subject matter. We decided to use these new articles as supplementary materials but not integrate them into the synthesis of the current review.

Results

This paper is a part of a larger study, and we have presented results related to study characteristics in a previous publication [3]. Multimedia Appendix 4 provides detailed information regarding the characteristics of the included reviews, including research questions or aims, review types, analysis methods, number of included studies, target populations, health issues, and reported digital health interventions in each review. In the current paper, to achieve our research objectives, we identified reviews that reported different intervention maturity stages, timing of the evaluation, and timing of the data collection. Additionally, we identified a set of evaluation indicators of the digital PEx and classified them into the three pre-defined categories (i.e., intervention outputs, patient outcomes, and healthcare system impact) which in turn included nine themes, and twenty-two subthemes. Furthermore, we highlighted evaluation approaches in terms of evaluation theories, study designs, data collection methods and instruments, and data analysis methods. We discovered that it was valuable to compare the evaluation objectives of the included studies. Therefore, we captured five typical evaluation objectives and the stakeholders involved, which clarified why and for who DH evaluators carried out the evaluation tasks. We present our detailed findings below.

Evaluation Objectives

Our review findings highlighted five typical evaluation objectives:

The first objective is to *broaden the general understanding of the digital PEx and guide evaluation research and practice* (n=11; 24.4%) [61-71]. For instance, one review [69] aimed to identify implications for future evaluation research and practice on mental health smartphone interventions by investigating UX evaluation approaches.

The second is to *improve the design, development, and implementation of the DHI in terms of a better digital PEx* (n=15; 33.3%) [61-70, 72-76]. As demonstrated in an included review [66], the evaluation of DHIs is critical to assess progress, identify problems and facilitate changes to improve health service delivery and achieve the desired outcomes.

The third is to *achieve evidence-based clinical usage and increase DHIs' adoption and uptake* (n=14; 31.1%) [61, 63, 64, 66-70, 72, 73, 75, 77-79].

The fourth is to *drive ongoing investment* (n=3; 6.7%) [61, 78, 79]: Without compelling economic supporting evidence, the proliferation of DHIs will not occur. Therefore, the sustained clinical use, successful implementation, and adoption of, and continued investment in DHIs requires more evaluative information to ensure resources are not wasted on ineffective interventions [61].

The fifth is to *inform health policy practice* (n=3; 6.7%) [63, 64, 69]. As two included articles stated [63, 64], ongoing evaluation and monitoring of DHIs is critical to inform health policy and practice. Additionally, in terms of the varied evaluation objectives, the evaluation activities serve different stakeholder groups, including program investigators, evaluators, and researchers; designers, developers, and implementers; end-users, patients, and healthcare providers; clients, investors; and governments, policy makers.

Evaluation Timing Considerations

Among the included studies, evaluations were carried out at various stages of the intervention to fulfill the five evaluation objectives. Our findings show that most reviews reported feasibility, efficacy, and pilot studies (n=32; [61, 62, 64, 66-72, 75, 77-97]), then investigated effectiveness (n=20; [61, 66, 68, 69, 71, 73, 77, 78, 80, 83, 87, 88, 90-92, 94, 98-101]), and implementation studies (n=20; [62, 64, 66, 69, 70, 72, 76, 78, 81, 83, 86, 89-91, 93, 95, 96, 98, 99, 102]). Notably, some reviews included more than one type of study. Our findings show the timing of evaluation can be directly at pre- and/or post-intervention [61, 64, 67, 68, 70-73, 77-80, 85, 86, 89, 90, 92, 93, 95, 96, 98, 99], from baseline to or at short- or long-term follow-up [66, 67, 70, 73, 75, 78-80, 84, 85, 87, 89, 90, 92, 93, 96, 99, 103, 104], during intervention use [84, 93], continued monitoring [64, 83], and even at dropout [61]. One study [92] suggested providing a period of technical training and conducting a baseline test to reduce evaluation bias caused by individual technology familiarity and novelty. As demonstrated by another study [61], pre- and post-intervention assessments using clinical trials can measure intervention effectiveness (e.g., patients' blood glucose levels). In terms of the timing of data collection, an included paper [61] suggested that evaluations directly after the intervention are appropriate so that the users retain fresh memories of the experience. To sustain intervention outcomes over a longer period of time, longitudinal evaluations and long-term follow up evaluations were recommended in two studies [63, 92].

Evaluation Indicators

Evaluation indicators relate to the goal to which the research project or commercial program intends to contribute. Indicators are defined as “a quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor” [105]. Based on our initial codes, we grouped the evaluation indicators into 3 main categories: intervention outputs, patient outcomes, and healthcare system impact. Each category contains several themes and subthemes (see Tables 4-6) and is discussed in detail below.

Intervention outputs

Intervention outputs are partially determined by the intervention inputs and processes (i.e., influencing factors and design considerations, such as personalized design [3]). We identified three themes and eight subthemes within this category (Table 2). The first theme, *functionality,* refers to the assessment of whether the DHIs work as intended. Subthemes include (1) the consistency of intended value (e.g., the ability of the DHIs to collect the amount of accurate clinical metrics in real time [64, 70, 82, 96]), (2) the quality of content and information (e.g., tailored content [64, 72, 84, 89, 91, 94, 97, 98]), (3) the appropriateness of intervention features (e.g., the degree of system setup [62, 77]), and (4) the usage of intervention theories (e.g., the presence of an underlying theoretical basis [62, 67, 76, 86, 91, 94, 96, 98, 100]). The second theme, u*sability,* refers to whether the DH system is used as intended [29]. Both technology quality attributes (e.g., the ease of use [61-64, 67, 69, 70, 75, 76, 79, 84, 94, 103]) and interaction design (e.g., intuitive interface design [75, 76, 102]) can be used for the usability evaluations. The third theme, *care quality,* refers to effective, safe, people-centered, timely, accessible, equitable, integrated, and efficient care services [106]. For example, the assessment of convenient care accessibility (e.g., care that fits into daily routines [61, 67, 70, 84, 85, 89, 94, 96] and the credibility of DHIs’ owners [61, 62]).

Table 4. Themes, subthemes, and evaluation indicators of the intervention outputs of the digital patient experience.

Themes and subthemes	Studies, n (%)	Evaluation indicators	References
Functionality n=36 (80%)			
Intended values	21 (46.7%)	• Ability to either change or maintain the user’s health state in a beneficial way: support self-management, shared decision-making, trigger actions, track and respond to changes. • Ability to collect clinical metrics: the number of monitored variables, the frequency, accuracy, concordance, timeliness, and visibility of monitoring.	[61, 62, 64, 69-71, 73-77, 82, 87, 89-91, 96, 99, 101, 104, 107]
Content and information	20 (44.4%)	• Quality of the content: evidence-based, tailored, relevance, practicality, consistency, clarity. • Amount of the information: comprehensible, completeness, glanceability (understandability), and conciseness. • Language of the information: simple nontechnical language, actionable message, a nonauthoritarian, friendly, and non-judgmental tone of voice.	[62-64, 67, 69, 72, 73, 75, 76, 79, 84, 88, 89, 91, 92, 94, 97, 98, 101, 103]
Intervention features	20 (44.4%)	• Appropriate features that meet the intended values: activity planning, activity scheduling,	[61-64, 68-73, 75-77, 84, 88,

		activity tracking, diary, alerts, journal, feedback, reminders. • Degree of setup, maintenance, and training: ready to use, initial training, ongoing education. • Channel or mode of delivery: phone calls, social media, mobile applications, web, video, devices, and wearable kit.	89, 91, 94, 97, 102]
Theory-based interventions	11 (24.4%)	• Presence or absence of an underlying theoretical basis: behavior change theory, social presence, a quality certification.	[62, 67, 76, 79, 86, 91, 94, 96-98, 100]
Usability n=26 (57.8%)			
Technology quality attributes	24 (53.3%)	• Technology operability: the ease of use, learnability, memorability, readability, efficiency, system errors, product, or service. • Technology standards and specifications: interoperability, integration, scalability, ergonomics, connectivity, adaptability, flexibility, accuracy, and reliability.	[61-65, 67-70, 72, 73, 75, 76, 79, 81, 83, 84, 90, 94, 96, 99, 101, 103, 104]
Interaction design	17 (37.8%)	• Use of human-centered design methodologies during the development process: co-design, user-centered design, inclusive design. • Design quality of system architecture, layout, and interface: intuitive, interactive, personalized, esthetic.	[61-65, 67-70, 75, 76, 79, 81, 91, 94, 96, 102]
Care quality n=30 (66.7%)			
Accessible care	27 (60%)	• Accessibility of care services: data, information, and HCPs. • Involvement of related stakeholders: family, friends, and peer-to-peer communication. • Accessibility to high quality care: timely, integrated, continuous, improved (more predictable daily life), convenient (fits into daily routines), and personalized care.	[61-64, 67, 70-72, 74-77, 79, 83-89, 92, 94, 96, 99, 100, 103, 107]
Safe and credible care	14 (31.1%)	• Credibility and accountability of care: the owners' credibility, third party verification. • Security of care: the number of medical errors. • Privacy of care: the presence of general privacy notifications, the documentation of individual access to user private data, regulation compliance.	[61-64, 75, 76, 79-81, 87, 88, 100-102]

Patient Outcomes

Studies used a variety of quantitative and qualitative factors and variables to measure and describe patient outcomes (Table 3), referring to five themes (emotional outcomes, perceptual outcomes, capability outcomes, behavioral outcomes, and clinical outcomes), and twelve subthemes. *Emotional outcomes* are about patients' positive or negative feelings that result from the use or anticipated use of DHIs. For example, a high level of patient satisfaction [61, 63, 65, 69-75, 77, 80, 82-84, 90, 92, 97-99, 102-104] is a typical positive feeling. Increased concerns about data privacy and security [63, 72, 75, 79, 84, 94, 103, 107] is a frequently mentioned negative feeling. *Perceptual outcomes* are the informed states of mind or non-emotional feelings the patients achieve before, during or after using the DHIs [77], including patients' initial attitudes toward the DHIs (e.g., internal motivation [61, 72, 77, 83, 85, 86, 96]), patient-to-provider relationships, e.g., those that are enhanced by perceived improved accessibility to HCPs [61, 75, 77, 79, 83, 84, 86, 96, 100] versus those that are interfered with by perceived loss of face-to-face contacts [69, 72, 74, 79, 84, 85, 107]), perceived empowerment (e.g., increased confidence in managing their own health conditions [64, 77, 83, 85, 86, 98]) and burden (e.g., increased perception of restriction [63, 84-86, 93, 100, 103, 107]), and overall acceptance of the DHIs (e.g., willingness to use [69, 70, 75, 80]). *Capability outcomes* refer to the improvement of patients' self-management autonomy, health knowledge, and clinical awareness. DHIs may be effective at improving their independency, self-management autonomy, problem-solving and decision-making skills [61, 70, 72, 73, 77, 79, 83-87, 92, 96, 100, 103]; gaining health literacy, knowledge or understanding of their health conditions or care plans [61, 64, 80, 83, 87, 96, 98, 100, 107]; and raising their clinical awareness to be more certain of when it was necessary to seek medical attention [77, 79, 80, 86, 100]. *Behavioral outcomes* entail activities that the patients adopt due to DHIs [77], including adherence to the intervention (e.g., dropout rates [69, 73, 77, 82, 84, 90, 92]), self-management behaviors (e.g., physical and diet activities [73, 75, 82, 86, 92, 96, 97]), and patient-to-provider communication (e.g., increased interactions between patients and HCPs [67, 71, 72, 77, 79, 83, 85, 87, 89, 96, 100]). *Clinical outcomes* are about individual health conditions, the main intentions of DHIs. For example, a reduction of anxiety, depression, and stress [67, 69-73, 77, 79, 83, 84, 86, 89, 90, 97, 103] and increased symptom control [75, 77, 79, 83, 86, 96-98, 104] can help to measure the individual health conditions.

Table 5. Themes, subthemes, and evaluation indicators of patient outcomes of the digital patient experience.

Themes and subthemes	Studies, n (%)	Evaluation indicators	References
Emotional outcomes n= 32 (71.1%)			
Positive emotions	31 (68.9%)	• Patient satisfaction • A sense of reassurance • Well-being • A sense of security • Peace of mind • A sense of belonging	[61, 63, 65, 69-75, 77, 79, 80, 82-86, 89, 90, 92, 93, 96-100, 102-104, 107]
Negative emotions	16 (35.6%)	• Concerns • Fears • A sense of uncertainties • Dissatisfaction • A sense of frustration • A sense of insecurity • Worries	[63, 69, 70, 72, 74, 75, 79, 84, 86, 93, 94, 96, 99, 100, 103, 107]
Perceptual outcomes n=32 (71.1%)			
Empowerment	23 (51.1%)	• Perceived values • Quality of life • Confidence • Self-efficacy • Comfort	[61, 63, 64, 69, 70, 73-75, 77, 79, 82-86, 92, 93, 96-100, 103]
Acceptability	19 (42.2%)	• Degree to which technology, treatment, and care services are accepted: willingness to use, intention to use, intention to continue using, likelihood to recommend	[61, 64, 67-70, 73, 75, 77, 80, 82-84, 86, 88, 90, 94, 99, 103]
Connectedness	16 (35.6%)	• Relationships between patient and provider: closeness, detachment, trust, or doubts	[61, 69, 72, 74-77, 79, 80, 83-86, 96, 100,

			107]
Attitudes	14 (31.1%)	• Initial beliefs, preferences, and expectations • Impression of the excellence of the DHIs • Interpretation of the DHIs • Motivation to change behavior	[61, 63, 72, 73, 75, 77, 80, 82-86, 88, 96]
Burden	12 (26.7%)	• Perceived burden and restriction • Discomfort • Unconfident	[63, 74, 75, 80, 84-86, 93, 97, 100, 103, 107]
Capability outcomes n=19 (42.2%)			
Autonomy and knowledge-gaining	19 (42.2%)	• Participants' level of informed state of mind after using the DHIs: clinical awareness • Patients' level of health knowledge: health literacy, skills, and understanding • Patients' ability to make clinical decisions: problem solving, shared decision-making	[61, 64, 70, 72, 73, 77, 79, 80, 83-87, 92, 96, 98, 100, 103, 107]
Behavioral outcomes n=26 (57.8%)			
Adherence	19 (42.2%)	• Initial, sustained use of certain features • Download and deletion rates • Completion rates • Dropout rates • Speed of task completion	[63, 69-71, 73, 75, 77, 82-84, 86, 87, 90, 92, 96-99, 103]
Self-management behaviors	17 (37.8%)	• Number of individuals exercising regularly, and/or using dietary behaviors compared to the total participants • Engagement of treatment, self-care, and help-seeking behavior	[61, 67, 69, 70, 73, 75, 77, 82, 83, 86, 89, 92, 93, 96-98, 100]
Patient-provider communication	11 (24.4%)	• Number and frequency of patient-provider contacts • Engagement of patient-provider communication • Quality of patient-provider communication (e.g., % patients reporting HCPs communicated well)	[67, 71, 72, 77, 79, 83, 85, 87, 89, 96, 100]
Clinical outcomes n=23 (51.1%)			
Health conditions	23 (51.1%)	• Level of pain and symptoms control • Status of physical health • Level of health or treatment-related anxiety, depression, and stress • Mortality rates • Morbidity rates • Adverse effects	[67-73, 75, 77, 79, 82-84, 86, 87, 89, 90, 96-99, 103, 104]

Healthcare system impact

Healthcare system impact contains one theme and two subthemes. *Economic outcomes* refer to the cost-effectiveness and healthcare services utilization. In terms of cost-effectiveness, for example, studies report less out-of-pocket expenses for patients due to reduced care and travel costs [64, 71, 72, 87, 89, 96, 98, 99, 107], and greater time-efficiency due to shorter waiting, travel, and consultation time [74, 80, 84, 85, 89, 99, 104]. Furthermore, indicators related to healthcare service utilization, such as the reduced number of hospital [70, 72, 84, 98, 99] and emergency department [98, 99] visits can be used to assess savings regarding healthcare services .

Table 6. Themes, subthemes, and evaluation indicators of healthcare system impact of the digital

patient experience.

Themes and subthemes	Studies, n (%)	Evaluation indicators	References
Economic outcomes n=16 (35.6%)			
Cost-effectiveness	14 (31.1%)	- Out-of-pocket expenses for patients: care costs, travel costs - Time-efficiency of using the DHIs: waiting time, travel time, and consultation time - Reduction in overuse of services: printed materials	[64, 71, 72, 74, 80, 84, 85, 87, 89, 96, 98, 99, 104, 107]
Healthcare service utilization	8 (17.8%)	- Duration of consultations - Number of hospitals, primary care, and emergency department visits - Hospital admissions - Hospitalization - Proportion of referrals	[70, 72, 84, 89, 96, 98-100]

Evaluation Approaches

In addition to evaluation timing considerations and indicators, strategies, and specifically designed tools for collecting and analyzing data are required to set up the evaluation plan. Varied evaluation approaches were identified based on our initial codes; these are depicted in three aspects (see table 7-9): study designs, data collection methods and instruments, and data analysis approaches. Furthermore, we collected data related to evaluation theories that were used to guide study designs, data collection and analysis.

▪ *Evaluation theories*

Our findings show that in some cases, theories are used to guide the evaluation process. An included review [66] mapped varied DHI evaluation frameworks and models into conceptual, results and logical frameworks, as well as theory of change. Among the included reviews, national quality forum [71, 87], user experience model [101], American Psychiatric Association app evaluation model [69], Markov model [96], and consolidated framework for implementation research [64] were mentioned as evaluation frameworks or models for setting up, conducting, or analyzing the evaluation activities. In addition, theories from other fields such as frameworks or models related to *healthcare* (e.g., diabetes theory [64, 77], triple aims framework [99], chronic disease management model [66]), *behaviors* (e.g., social cognitive theory [67, 90, 101], behavior change theory [66, 67, 98]), *design* (e.g., human factors principles [95], inclusive design [65]), and *technology* (e.g., the unified theory of acceptance and use of technology [65, 72], Health Information Technology Usability Evaluation Model [75]) can be adopted to assess specific outputs, outcomes or impact. For example, behavior change theory can be used to guide the evaluation of patient behavioral outcomes [67].

▪ *Study designs*

The terminologies used to describe study designs were mixed in terms of different classification bases. Following Kumar Ranjit's work on research methodology [56], we identified four standards for classifying study designs in DH: the perspectives of mode of enquiry, nature of the investigation, reference period, and number of contacts with the study population. From the perspectives of "*mode of enquiry*", we found three types of study. The first is qualitative study design, such as phenomenology or ethnography studies. Second is quantitative study. The third is mixed-methods research and multiple methods research (i.e., more than one qualitative or quantitative method, such as using both focus groups and interviews to collect data). In addition, on the basis of *the nature of the investigation*, the collected primary studies among the included reviews were reported as observational studies versus experimental studies (randomized controlled trials and non-randomized trials); descriptive studies (case reports, case series, and cross-sectional) versus analytical studies (case-control or cohort studies). Based on the *number of contacts* with the study population, cross-sectional, before-and-after, and longitudinal studies were mentioned. Furthermore, in terms of the *reference period* (the timeframe in which a study explores a phenomenon, situation, event or problem), some included prospective designs while others reported retrospective study designs. Additionally, we note that others reported study designs from a design perspective, such as user studies, participatory design or contextual design, and design sessions.

Table 7. Study designs of evaluating the digital patient experience.

Study designs	Studies, n (%)	References
Mode of enquiry n=36 (80%)		
Qualitative study design - Phenomenology. - Ethnography.	35 (77.8%)	[61, 63, 64, 66, 67, 69, 70, 72-80, 82-91, 93, 94, 96, 98-100, 103, 104, 107]
Quantitative study	21 (46.7%)	[61, 63, 66, 69, 70, 72, 74-76, 78, 79, 82, 84, 91, 93, 94, 96, 98, 99, 103, 104]
Mixed-methods research (and multiple methods research)	17 (37.8%)	[61, 63, 64, 69-71, 74, 76, 79, 84, 86, 88, 93, 94, 96, 99, 103]
Nature of the investigation n=33 (73.3%)		
Experimental studies - Randomized controlled trials. - Non-randomized trials.	25 (55.6%)	[61, 66-68, 70-72, 78-80, 83, 84, 86-93, 96-99, 103, 104]
Observational studies	9 (20%)	[68, 80, 84, 88, 92-94, 96, 99]
Descriptive studies - Case reports. - Case series. - Cross-sectional.	7 (15.6%)	[63, 64, 76, 79, 80, 82, 99]
Analytical studies - Case-control. - Cohort studies.	6 (13.3%)	[63, 68, 79, 96, 99, 102]
Number of contacts n=21 (46.7%)		
Cross-sectional	8 (17.8%)	[63, 64, 76, 80, 82, 99]
Longitudinal studies	6 (13.3%)	[63, 70, 79, 101, 103, 107]
Before-and-after	4 (8.9%)	[61, 67, 68, 70-73, 79, 80, 89, 90, 95, 98]
Reference period n=10 (22.2%)		
Prospective designs	8 (17.8%)	[68, 70, 79, 80, 89, 97, 99, 102]
Retrospective study designs	4 (8.9%)	[64, 68, 99, 103]
The design study n=4 (8.9%)		
User research study	3 (6.7%)	[63, 68, 95]
Participatory design or contextual design	1 (2.2%)	[77]
Design sessions	1 (2.2%)	[63]

Data collection methods and instruments

Varied data collection methods were used among the included reviews: questionnaires, surveys, interviews, focus groups, observations, log data, open-ended questions, Likert scales, usability testing, diaries, contextual inquiry, needs assessment, performance tests, field notes, workshops, forms, think-aloud, benchmark testing, human impact assessment methodologies, and personas. Notably, these data collection techniques appeared as a mixed combination in some studies. Additionally, we found varied standard evaluation tools and performance tests used to collect the digital PEx-related data in 18 of the included papers [61, 63, 65, 67, 69, 73, 75, 79, 83, 84, 87, 90, 92, 95, 97, 99, 102, 104] examples are: System Usability Scale (SUS) [61, 69, 90]; Patient Activation Measure (PAM) [83, 92]; Patient Health Questionnaire-9 (PHQ-9) [83, 97]; and Beck Depression Inventory (BDI) [83, 97]. However, none of these tools is designed for evaluating the digital PEx; most are designed or modified to evaluate UX, PEx in general, or the usability of specific DHIs.

Table 8. Data collection methods of evaluating the digital patient experience.

Data collection methods	Studies, n (%)	References
Questionnaires	33 (73.3%)	[61, 63, 64, 66, 67, 69, 70, 72-77, 79, 80, 83, 84, 87-95, 97-101, 103]
Surveys	32 (71.1%)	[61, 63, 66, 67, 69-77, 79-82, 84, 85, 87-91, 94, 95, 97-99, 101-103]
Interviews	31 (68.9%)	[61, 63, 64, 67-70, 73-77, 79-81, 83-86, 88-95, 99, 100, 103, 107]
Focus groups	19 (42.2%)	[63, 64, 66-68, 71, 74-76, 84, 86, 88, 89, 91, 93-95, 103, 107]
Observations	17 (37.8%)	[63, 68, 74, 75, 77, 79, 80, 84, 86, 88, 92-96, 99, 107]
Log data	13 (28.9%)	[63, 69, 77, 79-82, 89, 91, 92, 98, 103, 107]
Open-ended questions	10 (22.2%)	[61, 64, 67, 70, 73, 75, 83, 85, 88, 94]
Likert scales	10 (22.2%)	[61, 66, 73, 75, 79, 90, 92, 97, 99, 101]
Usability testing	8 (17.8%)	[61, 65, 68, 72, 75, 89-91]
Diaries	6 (13.3%)	[61, 63, 76, 88, 98, 107]
Contextual inquiry	5 (11.1%)	[61, 64, 77, 88, 95]
Needs assessment	5 (11.1%)	[61, 85, 90, 91, 95]
Performance tests	5 (11.1%)	[68, 69, 73, 80, 92]
Field notes	4 (8.9%)	[64, 77, 93, 107]
Workshops	4 (8.9%)	[75, 76, 90, 91]
Forms	3 (6.7%)	[61, 80, 90]
Think-aloud	3 (6.7%)	[61, 76, 77]
Benchmark testing	2 (4.4%)	[69, 95]
Human impact assessment methodologies	1 (2.2%)	[103]
Personas	1 (2.2%)	[95]

- ### *Data analysis approaches*

Our findings showed that different types of data were used to evaluate the digital PEx, for example, self-reported data [82] and observable or monitored data [69]. To analyze the evaluative information, varied data analysis methods were reported among the included reviews, including statistical analysis, thematic analysis, content analysis, grounded theory, framework analysis, heuristic analysis, cost analysis, task analysis, text analysis, document analysis, failure analysis, inductive analysis, deductive analysis, formal analysis, and decision analytic approach.

Table 9. Data analysis approaches of evaluating the digital patient experience.

Data analysis approaches	Studies, n (%)	References
Statistical analysis	15 (33.3%)	[67-69, 73, 78-80, 82-84, 90, 92, 98, 99, 104]
Thematic analysis	11 (24.4%)	[64, 69, 77, 84, 85, 88, 93, 96, 100, 103, 107]

Content analysis	9 (20%)	[61, 64, 71, 84, 85, 88, 94, 100, 107]
Grounded theory	7(15.6%)	[61, 64, 69, 88, 93, 100, 107]
Framework analysis	5 (11.1%)	[64, 88, 93, 100, 107]
Heuristic analysis	4 (8.9%)	[69, 75, 88, 95]
Cost analysis	4 (8.9%)	[71, 78, 96, 99]
Task analysis	3 (6.7%)	[69, 91, 95]
Text analysis	2 (4.4%)	[74, 100]
Document analysis	2 (4.4%)	[79, 88]
Failure analysis	2 (4.4%)	[91, 95]
Inductive analysis	2 (4.4%)	[64, 107]
Deductive analysis	1 (2.2%)	[64]
Formal analysis	1 (2.2%)	[81]
Decision analytic approach	1 (2.2%)	[99]

Discussion

The goals of this umbrella review were to systematically review the evaluation timing considerations, indicators, and approaches of the digital PEx. Furthermore, we were able to identify five typical evaluation objectives and related audiences. The timing of a digital PEx evaluation should be a critical consideration when conducting an evaluation study, however, we found limited information about when to measure the digital PEx. Moreover, the identified evaluation indicators are often heterogeneous and appear related to different aspects of the digital PEx. In terms of evaluation approaches, various theories were reported among the included papers. Furthermore, we noted that not only did the evaluation methods differ between the reviews, but also the classification bases or perspectives used to describe these methods. Following our findings on when to measure, what to measure, and how to measure the digital PEx, we generated a step-by-step evaluation guide, and propose six research directions for the future studies.

When to measure?

DHIs alter throughout the product lifecycle, so to provide better quality results and evidence-based health practice, evaluations need to be incorporated into the intervention maturity stages [63, 64, 66, 88]. Our findings show that many studies were not performed in a real-world setting for a long period, most were either a feasibility or pilot study; these results are directly in line with previous findings [64, 73, 81, 90, 93, 97, 107]. Pilot or feasibility studies can help improve new intervention development, but only provide limited evidence for increasing sustained clinical usage and large-scale practice [66, 88]. Two studies [63, 85] reported there was a lack of information on the long-term experience. Others have shown that some solutions may be less sustainable outside of a trial context [88, 93]. Additionally, it is possible that participants were more adherent during the study period, and decreased their use of the apps over time [82]. Therefore, some authors call for further research on the digital PEx when incorporating the DHIs into existing health care services and processes [84]; there is a need to move DHIs from promise into policy and practice [64].

One study [80] reported significantly different evaluation results before and after the treatment. It is likely that patients' initial emotional state or understandings of DHIs may impact their final dPEx evaluation outcomes. A pre-test on individual difference would therefore be a valuable step to limit evaluation bias, as noted in a study [92]. We found the data gathered can occur at a specific moment or at different time points along the care pathway to reflect a rapid or delayed digital PEx. Thus, posttreatment evaluations should account for the recall bias caused by the time delay between treatment and recollection of experience, as has been noted in previous studies [61, 84]. In line with other studies [29, 72], we believe that real-world testing and direct feedback from actual users will help improve the usability of DHIs and directly benefit new users.

What to measure?

In comparison to intervention outputs and healthcare system impact, we discovered more evaluation indicators related to patient outcomes. We assume that this is due to the consideration of the strength of the evidence and the study's duration. Patient outcomes enable the identification of patients' actual experiences and reactions in uncontrolled settings, providing evidence for clinical usage and further improvements. However, intervention outputs seem more suitable for exploring experts (e.g., designers, healthcare professionals, policy makers) or patients' anticipated understandings of DHIs in the early stages of design, and for addressing any potential system barriers. The healthcare system impact can be useful in predicting the sustainability of the DHIs on a large scale through a long-term study.

We used a set of themes and subthemes to describe each category. For instance, patient outcomes contain emotional, perceptual, capability, behavioral, and clinical outcomes as noted in two studies [68, 83]: one categorized the variables of patient engagement as behavioral, cognitive, and emotional outcomes; the other used biomarkers, perceptions, and behaviors to describe patient clinical outcomes regarding DHIs. Furthermore, we note that the evaluation outcome of one indicator is often unable to anticipate the outcome of another indicator. For instance, some patients reported high acceptance to a certain DHI but they rarely used it [103]. Aligned with the arguments among the differences between patient satisfaction, PEx, PREMs, and PROMs [35, 37, 47, 48], our findings indicate that digital PEx evaluations are not equivalent to the measurement of patient satisfaction, PEx, PREMs, or PROMs, but that these measures can be used to assess some of the digital PEx. We show that the priorities of evaluation indicators can differ between projects. In terms of what to measure first, as stated in a study [108], the goal of evaluations should be to focus on those processes that should be optimized by the digital catalyst. Furthermore, the evaluation indicators need be continually updated as the DH landscape is rapidly evolving and the technology infrastructure is constantly shifting [62].

How to measure?

As demonstrated in an included review [66], an evidence-based theoretical evaluation framework is helpful in informing the evaluation process. Across the included reviews, we found that not only specifically designed evaluation theories were used to guide the evaluation activities, but also theories from other fields were adopted to assess the evaluative data. We found varied traditional approaches across the included reviews. In addition, our results showed that more than half of the included reviews reported RCTs in their studies. RCTs were recommended in two reviews [109, 110] to evaluate DHIs for stronger evidence. However, a recent systematic review [111] noted that only a handful of clinical decision support systems have been tested in this way. Others argued that there is a tension between the amount of time needed for evidence generation with traditional approaches and the speed of digital product development and iterative upgrading [16, 18], which requires more innovative ways for fast evidence generation [18].

We identified a wide range of evaluation methods and instruments, while most were modified based on the evaluations for traditional face-to-face treatment or usability testing in human-computer interactions. This is also in line with findings from previous studies [61, 69, 80, 87]. Semi-structured interviews and questionnaires were the most common evaluation methods for collecting evaluative data among the included reviews, this is in line with previous studies [61, 84]. Semi-structured interviews are the key methods used to understand the details of UX [67, 69, 70, 73, 74], while questionnaires are often modified from existing assessments to assess larger scale interventions [61, 69]. It is likely that more in-depth, observational data collection methods are necessary to better capture experience data [61, 74]. The use of a descriptive approach might be appropriate for a smaller sample size, collecting qualitative data through surveys, focus groups, and interviews [84]. Standard functional questionnaires may be preferred when DHIs are compared with other interventions [61]. However, we found the detailed interview outlines or questionnaires were generally not published, as has been mentioned in another study [61]. Comprehensive information on user evaluation methods and results is often lacking [73]. The determination of evaluation approaches depends on the specific context. In alignment with two studies [4, 40], we state that the choice of evaluation approaches heavily depends on evaluation objectives, timing, indicators, and evaluation requirements and resources. An included review [66] recommends using multiple research methods, such as combining qualitative, quantitative, co-design principles, and process measures, for evaluation designs.

Thanks to the use of digital technologies [62, 83], patients' illness experience, and what they feel when taking part in a healthcare intervention can be monitored. However, we found that these may blur the boundaries between interventions, monitoring, and evaluations. For example, the diary function can be used as an intervention feature (e.g., a self-management diary to track symptoms and identify exacerbations [86]), as a monitoring tool (e.g., diary entries [107] or adherence [98]), or an evaluation method (e.g., to capture user feedback [61]). Furthermore, a study indicated that with the advancement of technology, the ability of DHIs to collect "passive data" for assessing the digital PEx may gain more attention and eventually eclipse the utility of DH-aided self-report [82]. Finally, we argue that not only does the design process need to involve multiple stakeholders, but this is also a requirement for the evaluation process. Both end-users and experts can contribute to the evaluation activities [61]. This aligns with a recent study which suggests that digital solution evaluation requires collective efforts from multiple parties, such as health authorities, healthcare providers, and manufacturers [18].

Design Implications

Our analysis shows that the evaluation of a DHI follows much the same evaluative process as that of traditional interventions, which supports a previous study [61]. To make evaluation findings more comparable, more rigorous studies and standardized evaluations are suggested, including unified terminology [61, 73, 76], predefined measurable indicators [87, 89], standardized methods [69, 74], validated instruments [92, 104], uniform time intervals [92], and adequate patient selection [89]. Intervention characteristics (e.g., aims, expected outcomes, elements, length, frequency, and duration), study designs (e.g., sample size, time period, regulations, investigator, evaluators, recruitment, ethics, topic guides, or questions asked by the researchers), objectively measured patient health outcomes, and adverse events should all be carefully considered when conducting and reporting an evaluation study [61, 68, 77, 92].

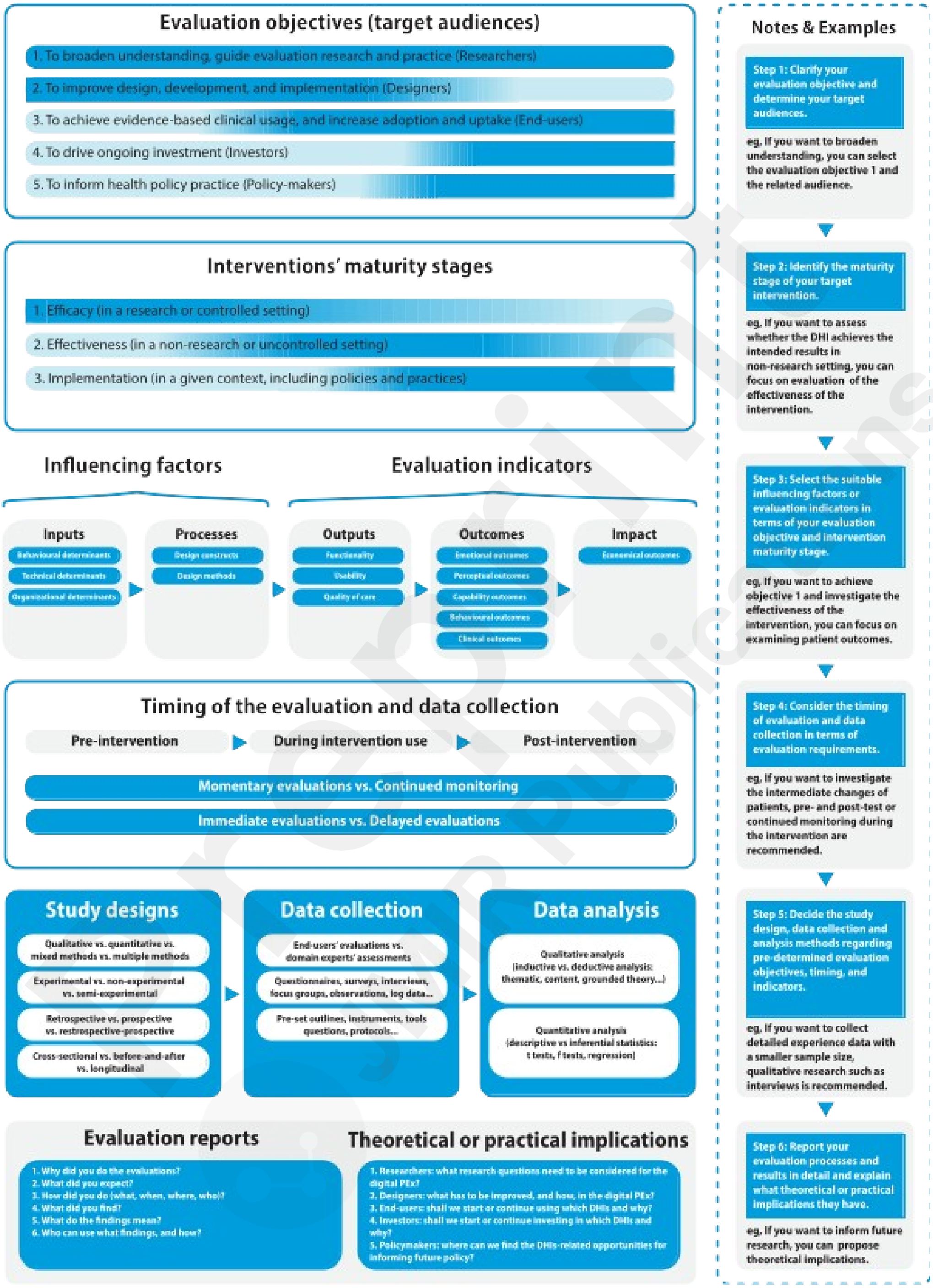
Inspired by the challenges for the evaluation of DHIs [18], shaped by the PRISM framework [55], the M&E Guide [29], PEx measures [37], and our previous publications on influencing factors and design considerations of digital PEx [3, 112], based on the findings in this current study, we have developed a step-by-step evaluation guide for DH innovators, such as designers, developers, and evaluators (Figure 2):

- The first step is to clarify evaluation objectives and determine the target audience(s) for the evaluation activity. We propose five typical evaluation purposes and their related audiences. The selection of evaluation objectives can somehow help determine at what stages to evaluate the DHI. For example, we consider effectiveness and implementation studies more appropriate for achieving evidence-based clinical usage and increasing adoption and uptake compared to efficacy studies.

- The second step is to determine the intervention contexts and foci in terms of the intervention maturity stages, including efficacy, effectiveness, and implementation. The determination of the evaluation stage is not only due to the evaluation objective but also to the current condition of the DHI. The determination of evaluation objectives and identification of the evaluation stage affect the consideration of influencing factors and evaluation indicators at the next step. For example, the evaluation of patient outcomes in an uncontrolled setting can provide evidence for clinical use and further improvement.
- The third step includes a set of influencing factors (i.e., inputs, processes,), and evaluation indicators (i.e., outputs, outcomes, and impacts) that can be used for further formulating evaluation constructs. The former is more appropriate for formative evaluations, which often happen during the design and development process, while the latter is suitable for summative evaluations, which often occur during and after the implementation process.
- In the fourth step, we show two types of evaluations. Based on the frequency of evaluations, we can capture momentary experiences prior to, during, and following an intervention or monitor continuous feedback throughout the intervention. Regarding the time interval between the intervention and evaluation, assessments can reflect immediate experiences directly after the intervention or recalled experiences over an extended period.
- In the fifth step, we present a variety of evaluation approaches that can be used to plan and carry out specific evaluation activities, such as study designs, data collection methods and instruments, and data analysis approaches. The consideration of study designs often affects the strength of the evidence and determines the data collection and analysis methods. Behavioral data may provide stronger evidence than opinion data. Qualitative methods, such as interviews, are more appropriate to collect in-depth experience data for a smaller sample size in the early intervention development stages; quantitative methods, such as questionnaires, are more suitable for investigating experience data at a large scale or comparing it with other interventions during or after implementation stages.
- In the sixth step, we propose six questions for the evaluation investigators to guide them in reporting the evaluation results and five questions to inspire them to generate theoretical or practical implications for responding to the related stakeholder groups. The answers to these eleven questions should reflect the evaluation processes and serve the evaluation objectives.

The guide can be used when setting up a digital PEx evaluation plan or guiding evaluation practice. Notably, the interrelationships between these six steps are not fixed; the whole evaluation plan is an iterative process; the decisions made at previous steps may influence the following steps, and vice versa. Additionally, other considerations beyond this guide can also impact the evaluation process, such as human, time, and financial resources. Our guide presents an ideal way to conduct the evaluation of the digital PEx; however, in the real world, the order of these steps may be changed or even some steps may be skipped depending on the specific project context. For instance, in certain assessment procedures, selecting an evaluation construct—such as usability—may come first rather than taking evaluation objectives or target audiences into account. We developed this guide based on our literature analysis. It provides an overview of the most common evaluation timing considerations, indicators, and approaches used to collect digital PEx-related data. However, it may be incomplete and require updating in the future. For example, due to the methodological limitations, we didn’t provide concrete recommendations on which evaluation approaches are superior for what types of DHIs. We believe that without providing a specific context and concrete project requirements, it is hard to draw a conclusion on this.

Figure 2. Digital patient experience evaluation guide



Limitations

This study has several limitations. First, we note possible resource restrictions and the newness of the field, which may have led to missing articles. To overcome this, we searched three databases and used the snowballing method. Besides, we performed an updated literature search to check if there were any meaningful new insights that would significantly change our conclusions. According to our knowledge, although there were some newly published reviews in this area, we confirmed that our results were quite stable, and the newly identified studies were unlikely to significantly impact our results. Second, we could not perform a quality assessment due to the diversity in reviews and methodological limitations. As previous studies on investigating umbrella reviews pointed out, there are currently no official standards for determining the certainty of evidence when performing umbrella reviews [113, 114]. Besides, among the included reviews, only 24 (53.3%) assessed the risk of bias and used diverse quality assessment instruments. After a thorough try, we found none of these instruments suitable for assessing the varied reviews included. They range from systematic reviews, scoping reviews, comprehensive overviews, and general literature reviews, which include diverse primary and secondary studies that go beyond randomized controlled trials (RCTs) and non-randomized studies of interventions, making it difficult to use a standardized way to assess the quality of evidence across the wide range of the included reviews. However, we tried to reduce the risk of bias by only including reviews that were published in peer-reviewed journals. Third, reviewing secondary research may lead to omitting crucial information and reporting bias. To minimize potential bias, we used the most common terms used across the included papers as themes and sub-themes. Due to the cross-disciplinary nature of the topic, there is a lack of consistency or clarity in the terminology used to describe the evaluation indicators and approaches. For instance, in one article [63], a user study is pitched at the same level as interviews or observations, while in another [64], interviews and questionnaires are methods that form part of “user study” research. Additionally, information related to evaluation approaches was reported at different levels among the included studies. For example, one study provided information related to data collection methods, such as focus groups, design sessions, and questionnaires [63]. While another study reported information related to study designs, including qualitative, quantitative, and mixed-methods designs [79]. These inconsistencies complicated comparison between different studies. To counter this, we analyzed the different classification bases behind these study designs. Last, we could not draw firm conclusions about which evaluation approaches are better suited for which types of digital health interventions. Due to the nature of this study being a review of reviews, details like the characteristics of DHIs are not always adequately covered in each included review. Additionally, the included reviews contain a large number of primary studies, which makes referring back to each primary study challenging. These limits the classification of DHIs in the current study. Besides, the included reviews represent a wide range of studies, making comparison across the included reviews challenging. Additionally, we decided this is out of scope in the current study. When we planned this study, we considered to first providing an overview of diversities instead of a best practice recommendation was a more suitable starting point for this new area. Consequently, we aimed to map possible evaluation considerations and approaches for evaluating the digital PEx instead of already starting to discuss which approach is better. However, we encourage future research to address this consideration.

Future research

Considering our research limitations, to further facilitate evaluations of the digital PEx, we propose six future research directions. First, further research into how one indicator mediates another indicator's impact on the digital PEx is needed. For example, is there a correlation between the clinical outcomes and the perceptual outcomes? To explore this, we performed an experimental study to investigate whether patients’ initial pain perception and technology acceptance (using virtual reality distraction) impacted their experienced pain during wound care treatment. Our findings will be published in another article. Second, the variables that influence the selection or prioritization of evaluation indicators and approaches should be further investigated. For example, it would be valuable to investigate whether some evaluation indicators and approaches are better suited for evaluating certain types of digital health interventions according to the strength of the evidence and the length of the evidence generation time. Third, agreement is needed on standardized measures to evaluate the digital PEx, especially innovative approaches for faster and high-quality evidence generation. In a follow-up interview study, we aim to summarize often-used agile evaluation approaches based on designers' experiences. Furthermore, in cases where an interview or questionnaire is used to collect evaluative information, we recommend reporting the detailed interview outlines or questionnaires together with the evaluation results. Fourth, research is needed into how the intervention maturity stages and timing of the evaluation of the evaluation affect the evaluation results. Fifth, future studies should not only investigate whether the DHIs achieve the intended results in a research setting, but also assess the long-term digital PEx regarding the uptake, institutionalization, and sustainability of evidence based DHIs in a given context and a real-world setting, including policies and practices. Lastly, research is needed on how to analyze and respond to the evaluative data. We recommend that future evaluation research and practice create theoretical and practical guidance on how to utilize the evaluative information.

Conclusion

To effectively improve the digital PEx, knowing how to evaluate the digital PEx is as important as knowing what factors influence the digital PEx and how to design the digital PEx. Evaluating the digital PEx requires clarifying the evaluation objectives, identifying stakeholder groups, considering the reasonable evaluation timings, choosing relevant evaluation indicators, and selecting appropriate evaluation approaches. Following our previous publication on the influencing factors and design considerations of digital PEx [3], we firstly identified five typical evaluation objectives and related stakeholder groups. We then described potential evaluation timing considerations in terms of four intervention maturity stages and three evaluation timings. We collected knowledge on evaluation indicators of the digital PEx and grouped them into three categories: intervention outputs, patient outcomes, and healthcare system impact. These were then classified in 9 themes (intervention functionality, usability, care quality, patient emotional outcomes, perceptual outcomes, capability outcomes, behavioral outcomes, clinical outcomes, and system financial outcomes) and 22 subthemes. Furthermore, we noted a set of common study designs, data collection methods and instruments, as well as data analysis methods, which can be used or adapted to evaluate the digital PEx. Based on our findings, we developed an evaluation guide to help DHI researchers, designers, and developers to further evaluate the digital PEx. Finally, we recommend six directions for further digital PEx evaluation research. Multimedia Appendix 5 provides more detail on the structure of this review.

Acknowledgments

This work was supported by the China Scholarship Council under Grant 201906790084.

Authors' Contributions

TW conceived, designed, led overall study conduct, carried out the data collection, led the analysis and interpretation of the data, and drafted the manuscript. TW and GG selected the papers. MM, RG, and GG participated in the overall study conduct and contributed to the analysis and interpretation of study data, and conceptualized, reviewed, and suggested modifications to the presentation of results. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Abbreviations

DH: digital health

DHIs: digital health interventions
eHealth: electronic health
HCPs: health care providers
ICT: information and communication technology
mHealth: mobile health
M&E: monitoring and evaluating digital health interventions
NQF: national quality forum
PEx: patient experience
PREMs: patient reported experience measures
PRISMA: the Preferred Reporting Items for Systematic Reviews and Meta-Analyses methodology
PRISM: the performance of routine information system management framework
PROMs: patient reported outcome measures
RCTs: randomized controlled trials
UX: user experience
WHO: world health organization

References

1. Jandoo T. WHO guidance for digital health: what it means for researchers. Digit Health. 2020;6:2055207619898984.
2. Kellermann AL, Jones SS. What it will take to achieve the as-yet-unfulfilled promises of health information technology. Health affairs. 2013;32(1):63-8.
3. Wang T, Giunti G, Melles M, Goossens R. Digital patient experience: umbrella systematic review. J Med Internet Res. 2022;24(8):e37952.
4. IQVIA Institute for Human Data Science Study: Impact of Digital Health Grows as Innovation, Evidence and Adoption of Mobile Health Apps Accelerate. IQVIA; 2017 [cited 2022 14, Jan]; Available from: https://www.iqvia.com/newsroom/2017/11/impact-of-digital-health-grows-as-innovation-evidence-and-adoption-of-mobile-health-apps-accelerate/.
5. Asadzadeh A, Kalankesh LR. A scope of mobile health solutions in COVID-19 pandemics. Informatics in Medicine Unlocked. 2021:100558.
6. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ digital medicine. 2020;3(1):14.
7. Alkire L, O'Connor GE, Myrden S, Köcher S. Patient experience in the digital age: An investigation into the effect of generational cohorts. Journal of Retailing and Consumer Services. 2020;57:102221.
8. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond “implementation”: digital health innovation and service design. NPJ digital medicine. 2018;1(1):48.
9. Constand MK, MacDermid JC, Bello-Haas D, Law M. Scoping review of patient-centered care approaches in healthcare. BMC Health Serv Res. 2014;14(1):1-9.
10. Philpot LM, Khokhar BA, DeZutter MA, Loftus CG, Stehr HI, Ramar P, et al. Creation of a patient-centered journey map to improve the patient experience: a mixed methods approach. Mayo Clinic Proceedings: Innovations, Quality & Outcomes. 2019;3(4):466-75.
11. Hollander J, Ward M, Alverson D, Bashshur R, Darkins A, DePhillips H, editors. Creating a framework to support measure development for telehealth. Washington, DC: National Quality Forum; 2017.
12. Whitten P, Love B. Patient and provider satisfaction with the use of telemedicine: overview and rationale for cautious enthusiasm. Journal of postgraduate medicine. 2005;51(4):294.
13. Riley PE, Fischer JL, Nagy RE, Watson NL, McCoul ED, Tolisano AM, et al. Patient and

provider satisfaction with telemedicine in otolaryngology. OTO open. 2021;5(1):2473974X20981838.
14. Altinisik Ergur G, Nuhoglu S, Cobanoglu C, Sengul M, Eryildiz N, Ergur A. The Patient Perspective of Telemedicine in the Context of COVID-19 Pandemic. Bulletin of Science, Technology & Society. 2022;42(1-2):39-53.
15. Keown OP, Parston G, Patel H, Rennie F, Saoud F, Al Kuwari H, et al. Lessons from eight countries on diffusing innovation in health care. Health Affairs. 2014;33(9):1516-22.
16. Desveaux L, Shaw J, Wallace R, Bhattacharyya O, Bhatia RS, Jamieson T. Examining tensions that affect the evaluation of technology in health care: considerations for system decision makers from the perspective of industry and evaluators. JMIR medical informatics. 2017;5(4):e8207.
17. World Health Organization. Classification of digital health interventions v1. 0: a shared language to describe the uses of digital technology for health. World Health Organization, 2018.
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 Digital Medicine. 2020 2020/08/27;3(1):110. doi: 10.1038/s41746-020-00314-2.
19. World Health Organization. Monitoring the implementation of digital health: an overview of selected national and international methodologies. Monitoring the implementation of digital health: an overview of selected national and international methodologies. 2022.
20. Kamillah S, Panduragan SL, Poddar S, Abdullah BF. Patients' Experiences in Using Diabetes Self-Management Application: A Scoping Review. Kesmas: Jurnal Kesehatan Masyarakat Nasional (National Public Health Journal). 2022;17(sp1).
21. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP open. 2020;4(1).
22. Research2Guidance. mHealth App Economics 2017/2018: Current Status and Future Trends in Mobile Health. 2017.
23. Koh A, Swanepoel DW, Ling A, Ho BL, Tan SY, Lim J. Digital health promotion: promise and peril. Health Promotion International. 2021;36(Supplement_1):i70-i80.
24. Eysenbach G, Powell J, Kuss O, Sa E-R. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. Jama. 2002;287(20):2691-700.
25. BINDHIM NF, NAICKER S, FREEMAN B, MCGEECHAN K, TREVENA L. Apps promoting illicit drugs—a need for tighter regulation? Journal of Consumer Health On the Internet. 2014;18(1):31-43.
26. Singh K, Diamantidis CJ, Ramani S, Bhavsar NA, Mara P, Warner J, et al. Patients' and Nephrologists' Evaluation of Patient-Facing Smartphone Apps for CKD. Clin J Am Soc Nephrol. 2019 Apr 5;14(4):523-9. PMID: 30898873. doi: 10.2215/cjn.10370818.
27. Fraser H, Bailey C, Sinha C, Mehl G, Labrique A. Call to action on global eHealth evaluation: consensus statement of the WHO Global eHealth Evaluation Meeting, Bellagio, September 2011. Bellagio, Italy: Bellagio eHealth Evaluation Group; 2011. Bellagio, Italy: Bellagio eHealth Evaluation Group. 2011.
28. World Health Organization. Digital implementation investment guide (DIIG): integrating digital interventions into health programmes. 2020 9240010564.
29. World Health Organization. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. 2016 9241511761.
30. Wolf PhD C, Jason A. Defining patient experience. Patient experience journal. 2014;1(1):7-19.
31. Jokela T, Iivari N, Matero J, Karukka M, editors. The standard of user-centered design and

the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. Proceedings of the Latin American conference on Human-computer interaction; 2003.
32. Tecco H. 2016 Year end funding report: A reality check for digital health. [cited 2022 May 17]; Available from: https://rockhealth.com/insights/2016-year-end-funding-report-a-reality-check-for-digital-health/.
33. World Health Organization. Global strategy on digital health 2020-2025. 2021.
34. Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? Social science & medicine. 2005;60(4):833-43.
35. LaVela SL, Gallan A. Evaluation and measurement of patient experience. Patient Experience Journal. 2014;1(1):28-36.
36. Member FCd. Timing of the evaluation and intervention cycle. 2022; Available from: https://europa.eu/capacity4dev/evaluation_guidelines/wiki/timing-evaluation-and-intervention-cycle-0#anchor2.
37. Coulter A, Fitzpatrick R, Cornwell J. Measures of patients' experience in hospital: purpose, methods and uses: King's Fund London; 2009. ISBN: 1857175859.
38. Bolton RN, McColl-Kennedy JR, Cheung L, Gallan A, Orsingher C, Witell L, et al. Customer experience challenges: bringing together digital, physical and social realms. Journal of Service Management. 2018.
39. Richardson B, Campbell-Yeo M, Smit M. Mobile application user experience checklist: A tool to assess attention to core UX principles. International Journal of Human–Computer Interaction. 2021:1-8.
40. Norman D, Nielsen J. The definition of user experience (UX). Nielsen Norman Group Publication. 2016;1:2.1.
41. Reeves R, Coulter A, Jenkinson C, Cartwright J, Bruster S, Richards N. Development and pilot testing of questionnaires for use in the acute NHS trust inpatient survey programme. Europe: Picker Institute. 2002.
42. Baker A. Crossing the quality chasm: a new health system for the 21st century: British Medical Journal Publishing Group; 2001. ISBN: 0959-8138.
43. Staniszewska S, Boardman F, Gunn L, Roberts J, Clay D, Seers K, et al. The Warwick Patient Experiences Framework: patient-based evidence in clinical guidelines. International Journal for Quality in Health Care. 2014;26(2):151-7.
44. NHS Patient Experience Framework. NHS National Quality Board; 2011 [cited 2021 May 07]; Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/215159/dh_132788.pdf.
45. Shandley LM, Hipp HS, Anderson-Bialis J, Anderson-Bialis D, Boulet SL, McKenzie LJ, et al. Patient-centered care: factors associated with reporting a positive experience at United States fertility clinics. Fertil Steril. 2020 Apr;113(4):797-810. PMID: 32147181. doi: 10.1016/j.fertnstert.2019.12.040.
46. Excellence NIfHaC. Evidence standards framework for digital health technologies. 2018 2022. Report No.: Contract No.: May 14.
47. Rockville M. What Is Patient Experience? : Agency for Healthcare Research and Quality; 2016 [cited 2021 May 10]; Available from: https://www.ahrq.gov/cahps/about-cahps/patient-experience/index.html.
48. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. Bja Education. 2017;17(4):137-44.
49. Hodgson P. User experience research and strength of evidence. 2017 [cited 2023 May 3]; Available from: https://userfocus.co.uk/articles/strength-of-evidence.html.
50. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA

statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of clinical epidemiology. 2009;62(10):e1-e34.
51. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health information & libraries journal. 2009;26(2):91-108.
52. Frank SR. Digital health care—the convergence of health care and the Internet. The Journal of ambulatory care management. 2000;23(2):8-17.
53. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative research in psychology. 2006;3(2):77-101.
54. Caulfield J. How to Do Thematic Analysis | Step-by-Step Guide & Examples. 2019 [cited 2023 May 4]; Available from: https://www.scribbr.com/methodology/thematic-analysis/.
55. Aqil A, Lippeveld T, Hozumi D. PRISM framework: a paradigm shift for designing, strengthening and evaluating routine health information systems. Health policy and planning. 2009;24(3):217-28.
56. Kumar R. Research methodology: A step-by-step guide for beginners: Sage; 2018. ISBN: 1526457083.
57. World health organization. Quality of care. [cited 2022 Feb. 10]; Available from: https://www.who.int/health-topics/quality-of-care#tab=tab_1.
58. Streefkerk R. Qualitative vs. Quantitative Research | Differences, Examples & Methods. [cited 2022 April, 20]; Available from: https://www.scribbr.com/methodology/qualitative-quantitative-research/.
59. Cumpston M, Chandler IV J. Chapter IV: Updating a review. Cochrane handbook for systematic reviews of interventions version. 2020;6.
60. Allen S. 2019 Global health care outlook: Shaping the future. 2019.
61. Feather JS, Howson M, Ritchie L, Carter PD, Parry DT, Koziol-McLain J. Evaluation methods for assessing users' psychological experiences of web-based psychosocial interventions: A systematic review. Journal of medical Internet research. 2016;18(6):e181.
62. Baumel A, Birnbaum ML, Sucala M. A systematic review and taxonomy of published quality criteria related to the evaluation of user-facing eHealth programs. Journal of medical systems. 2017;41(8):1-7.
63. Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM. Patient portal use and experience among older adults: systematic review. JMIR Medical Informatics. 2017;5(4):e38.
64. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: systematic review. Journal of medical Internet research. 2017;19(12):e8888.
65. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). International Journal of Medical Informatics. 2018;114:66-75.
66. Bashi N, Fatehi F, Mosadeghi-Nik M, Askari MS, Karunanithi M. Digital health interventions for chronic diseases: a scoping review of evaluation frameworks. BMJ health & care informatics. 2020;27(1).
67. Brigden A, Anderson E, Linney C, Morris R, Parslow R, Serafimova T, et al. Digital behavior change interventions for younger children with chronic health conditions: systematic review. Journal of medical Internet research. 2020;22(7):e16924.
68. Choi W, Wang S, Lee Y, Oh H, Zheng Z. A systematic review of mobile health technologies to support self-management of concurrent diabetes and hypertension. Journal of the American Medical Informatics Association. 2020;27(6):939-45.
69. Lemon C, Huckvale K, Carswell K, Torous J. A narrative review of methods for applying user experience in the design and assessment of mental health smartphone interventions.

International journal of technology assessment in health care. 2020;36(1):64-70.
70. Leonardsen A-CL, Hardeland C, Helgesen AK, Grøndahl VA. Patient experiences with technology enabled care across healthcare settings-a systematic review. BMC health services research. 2020;20(1):1-17.
71. O'Keefe M, White K, Jennings JAC. Asynchronous telepsychiatry: A systematic review. Journal of telemedicine and telecare. 2021;27(3):137-45.
72. Palacholla RS, Fischer N, Coleman A, Agboola S, Kirley K, Felsted J, et al. Provider-and patient-related barriers to and facilitators of digital health technology adoption for hypertension management: scoping review. JMIR Cardio. 2019;3(1):e11951.
73. Wesselman LM, Hooghiemstra AM, Schoonmade LJ, De Wit MC, Van Der Flier WM, Sikkes SA. Web-based multidomain lifestyle programs for brain health: comprehensive overview and meta-analysis. JMIR mental health. 2019;6(4):e12104.
74. Ingemann C, Hansen NF, Hansen NL, Jensen K, Larsen CVL, Chatwood S. Patient experience studies in the circumpolar region: a scoping review. BMJ open. 2020;10(10):e042973.
75. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M. Proposal for the User-Centered Design Approach for Health Apps Based on Successful Experiences: Integrative Review. JMIR mHealth and uHealth. 2020;8(4):e14376.
76. Wei Y, Zheng P, Deng H, Wang X, Li X, Fu H. Design Features for Improving Mobile Health Intervention User Engagement: Systematic Review and Thematic Analysis. Journal of Medical Internet Research. 2020;22(12):e21687.
77. Jalil S, Myers T, Atkinson I. A meta-synthesis of behavioral outcomes from telemedicine clinical trials for type 2 diabetes and the Clinical User-Experience Evaluation (CUE). Journal of medical systems. 2015;39(3):1-21.
78. Eze ND, Mateus C, Cravo Oliveira Hashiguchi T. Telemedicine in the OECD: An umbrella review of clinical and cost-effectiveness, patient experience and implementation. PloS One. 2020;15(8):e0237585.
79. Steindal SA, Nes AAG, Godskesen TE, Dihle A, Lind S, Winger A, et al. Patients' experiences of telehealth in palliative home care: scoping review. Journal of medical Internet research. 2020;22(5):e16218.
80. Swanepoel DW, Hall III JW. A systematic review of telehealth applications in audiology. Telemedicine and e-Health. 2010;16(2):181-200.
81. Memon M, Wagner SR, Pedersen CF, Beevi FHA, Hansen FO. Ambient assisted living healthcare frameworks, platforms, standards, and quality attributes. Sensors. 2014;14(3):4312-41.
82. Firth J, Torous J. Smartphone apps for schizophrenia: a systematic review. JMIR mHealth and uHealth. 2015;3(4):e4930.
83. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Frontiers in psychology. 2016;6:2013.
84. Jones L, Grech C. The patient experience of remote telemonitoring for heart failure in the rural setting: a literature review. Contemporary nurse. 2016;52(2-3):230-43.
85. Cox A, Lucas G, Marcu A, Piano M, Grosvenor W, Mold F, et al. Cancer survivors' experience with telehealth: a systematic review and thematic synthesis. Journal of Medical Internet Research. 2017;19(1):e11.
86. Morton K, Dennison L, May C, Murray E, Little P, McManus RJ, et al. Using digital interventions for self-management of chronic physical health conditions: a meta-ethnography review of published studies. Patient education and counseling. 2017;100(4):616-35.
87. Rising KL, Ward MM, Goldwater JC, Bhagianadh D, Hollander JE. Framework to advance

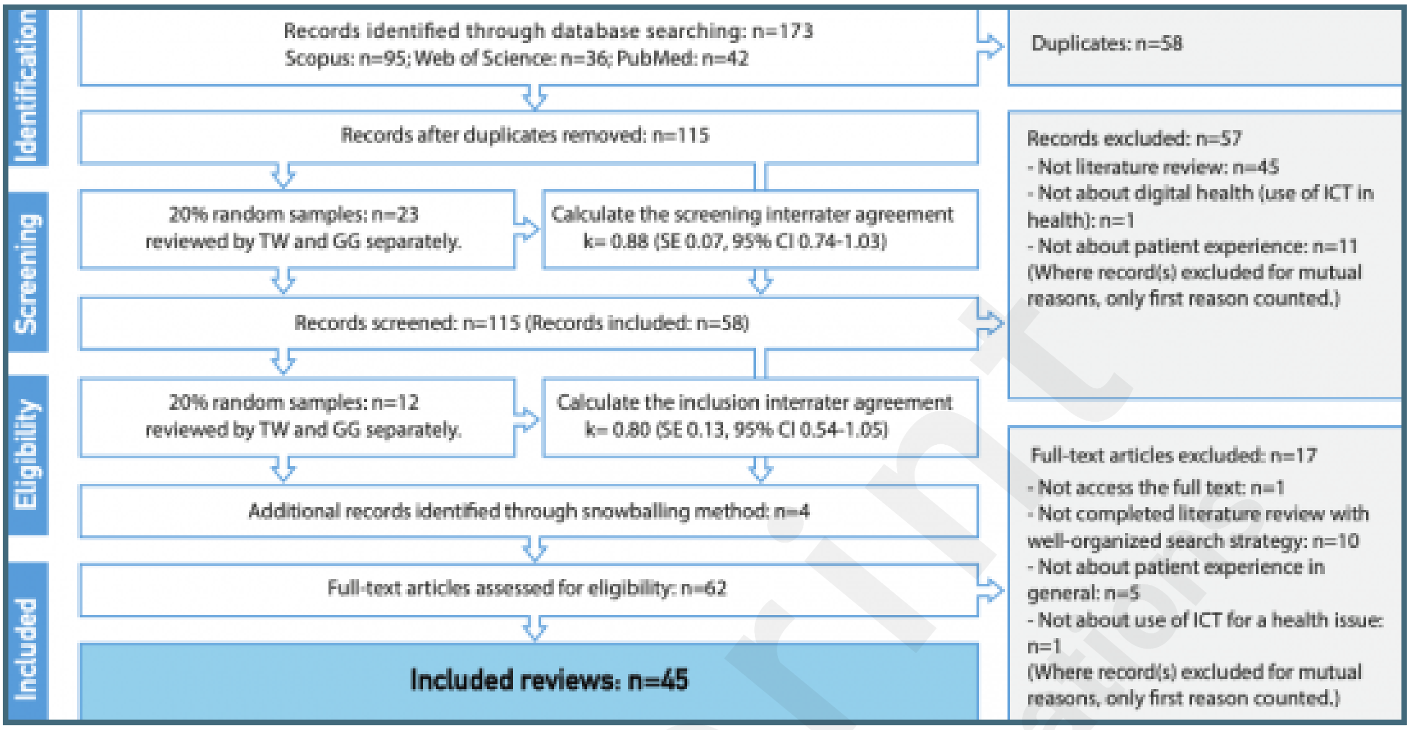
oncology-related telehealth. JCO clinical cancer informatics. 2018;2:1-11.
88. Ames HM, Glenton C, Lewin S, Tamrat T, Akama E, Leon N. Clients' perceptions and experiences of targeted digital communication accessible via mobile devices for reproductive, maternal, newborn, child, and adolescent health: a qualitative evidence synthesis. Cochrane Database of Systematic Reviews. 2019 (10).
89. De La Cruz Monroy MF, Mosahebi A. The use of smartphone applications (apps) for enhancing communication with surgical patients: a systematic review of the literature. Surgical innovation. 2019;26(2):244-59.
90. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. Journal of medical Internet research. 2019;21(7):e12869.
91. Søgaard Neilsen A, Wilson RL. Combining e-mental health intervention development with human computer interaction (HCI) design to enhance technology-facilitated recovery for people with depression and/or anxiety conditions: an integrative literature review. International journal of mental health nursing. 2019;28(1):22-39.
92. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res. 2013;15(2):e37.
93. Brunton L, Bower P, Sanders C. The contradictions of telehealth user experience in chronic obstructive pulmonary disease (COPD): a qualitative meta-synthesis. PLoS One. 2015;10(10):e0139561.
94. Lim S, Tan A, Madden S, Hill B. Health professionals' and postpartum women's perspectives on digital health interventions for lifestyle management in the postpartum period: a systematic review of qualitative studies. Frontiers in Endocrinology. 2019;10:767.
95. Fouquet SD, Miranda AT. Asking the Right Questions—Human Factors Considerations for Telemedicine Design. Current allergy and asthma reports. 2020;20(11):1-7.
96. Greenhalgh T, Shaw S. Understanding heart failure; explaining telehealth–a hermeneutic systematic review. BMC Cardiovascular Disorders. 2017;17(1):1-16.
97. Rincon E, Monteiro-Guerra F, Rivera-Romero O, Dorronzoro-Zubiete E, Sanchez-Bocanegra CL, Gabarron E. Mobile phone apps for quality of life and well-being assessment in breast and prostate cancer patients: systematic review. JMIR mHealth and uHealth. 2017;5(12):e8741.
98. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, MacKenzie AM, et al. Digital asthma self-management interventions: a systematic review. Journal of medical Internet research. 2014;16(2):e2814.
99. Liddy C, Drosinis P, Keely E. Electronic consultation systems: worldwide prevalence and their impact on patient care—a systematic review. Family practice. 2016;33(3):274-85.
100. Barken TL, Söderhamn U, Thygesen E. A sense of belonging: A meta-ethnography of the experience of patients with chronic obstructive pulmonary disease receiving care through telemedicine. Journal of advanced nursing. 2019;75(12):3219-30.
101. Cheung KL, Durusu D, Sui X, de Vries H. How recommender systems could support and enhance computer-tailored digital health programs: a scoping review. Digital Health. 2019;5:2055207618824727.
102. Werder M. Health information technology: A key ingredient of the patient experience. Patient Experience Journal. 2015;2(1):143-7.
103. Stokke R. The personal emergency response system as a technology innovation in primary health care services: an integrative review. Journal of medical Internet research. 2016;18(7):e187.

104. Chaudhry H, Nadeem S, Mundi R. How satisfied are patients and surgeons with telemedicine in orthopaedic care during the COVID-19 pandemic? A systematic review and meta-analysis. Clinical Orthopaedics and Related Research®. 2021;479(1):47-56.
105. World Health Organization. WHO evaluation practice handbook: World Health Organization; 2013. ISBN: 9241548681.
106. Quality of care. World Health Organization; [cited 2022 October 31]; Available from: https://www.who.int/health-topics/quality-of-care#tab=tab_1.
107. Walker RC, Tong A, Howard K, Palmer SC. Patient expectations and experiences of remote monitoring for chronic diseases: systematic review and thematic synthesis of qualitative studies. International journal of medical informatics. 2019;124:78-85.
108. Labrique A, Vasudevan L, Weiss W, Wilson K. Establishing standards to evaluate the impact of integrating digital health into health systems. Global Health: Science and Practice. 2018;6(Supplement 1):S5-S17.
109. Chung KC, Swanson JA, Schmitz D, Sullivan D, Rohrich RJ. Introducing evidence-based medicine to plastic and reconstructive surgery. Plastic and reconstructive surgery. 2009;123(4):1385.
110. Song JW, Chung KC. Observational studies: cohort and case-control studies. Plastic and reconstructive surgery. 2010;126(6):2234.
111. Pawloski PA, Brooks GA, Nielsen ME, Olson-Bullis BA. A systematic review of clinical decision support systems for clinical oncology practice. Journal of the National Comprehensive Cancer Network. 2019;17(4):331-8.
112. Wang T, Giunti G, Melles M, Goossens R. Design-Relevant Factors Affecting the Patient Experience in Digital Health: Preliminary Results of an Umbrella Systematic Review. Studies in health technology and informatics. 2022;290:862-6.
113. Sadoyu S, Tanni KA, Punrum N, Paengtrai S, Kategaew W, Promchit N, et al. Methodological approaches for assessing certainty of the evidence in umbrella reviews: A scoping review. Plos one. 2022;17(6):e0269009.
114. Choi GJ, Kang H. Introduction to Umbrella Reviews as a Useful Evidence-Based Practice. Journal of Lipid and Atherosclerosis. 2023;12(1):3.

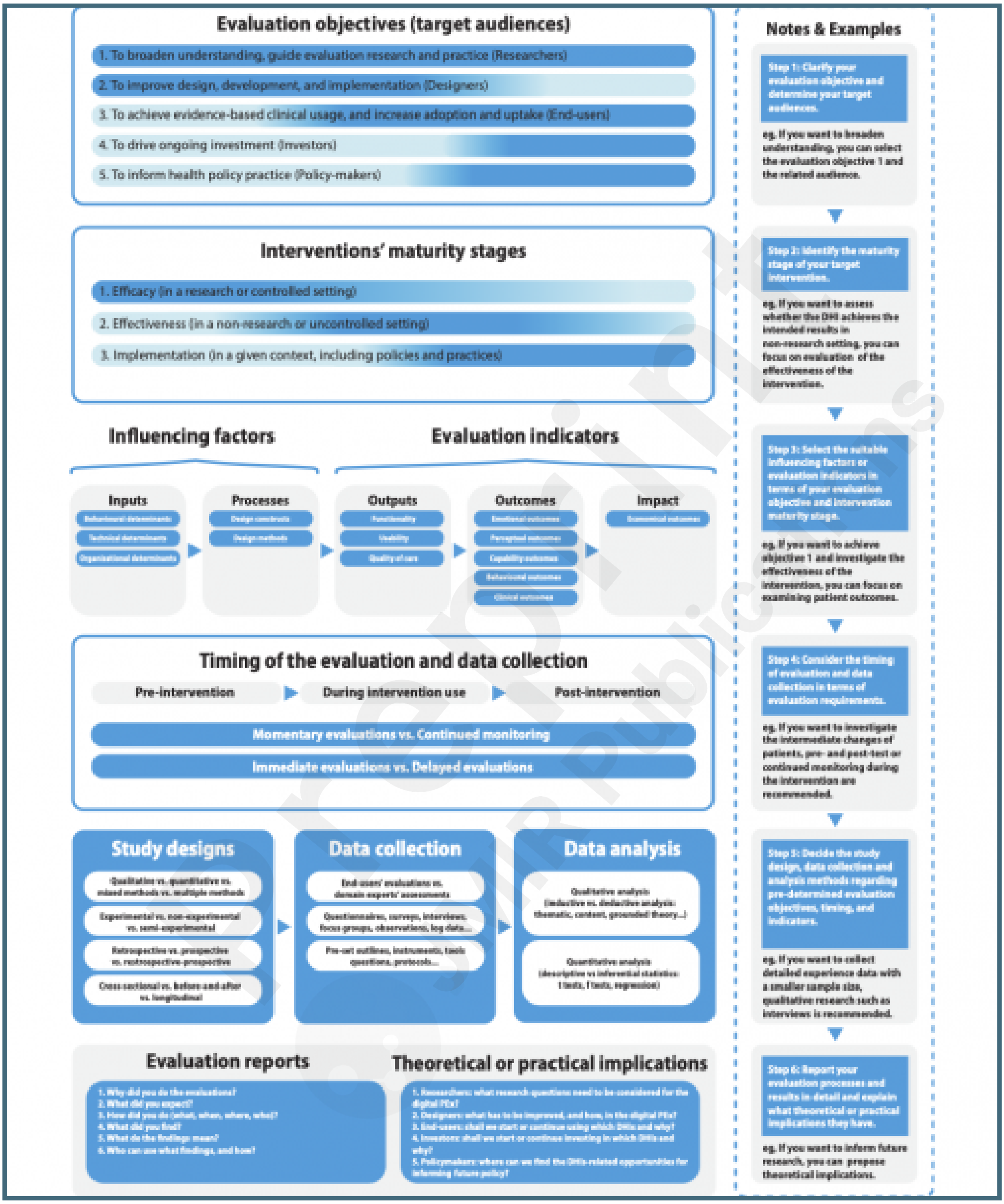
Supplementary Files

Figures

Study flow diagram.



Digital patient experience evaluation guide.



Evaluation reports

1. Why did you do the evaluations?
2. What did you expect?
3. How did you do (what, when, where, who)?
4. What did you find?
5. What do the findings mean?
6. Who can use what findings, and how?

Theoretical or practical implications

1. Researchers: what research questions need to be considered for the digital PEs?
2. Designers: what has to be improved, and how, in the digital PEs?
3. End-users: shall we start or continue using which DHIs and why?
4. Investors: shall we start or continue investing in which DHIs and why?
5. Policymakers: where can we find the DHIs-related opportunities for informing future policy?

Notes & Examples

Step 1: Clarify your evaluation objective and determine your target audiences.

eg. If you want to broaden understanding, you can select the evaluation objective 1 and the related audience.

Step 2: Identify the maturity stage of your target intervention.

eg. If you want to assess whether the DHI achieves the intended results in non-research setting, you can focus on evaluation of the effectiveness of the intervention.

Step 3: Select the suitable influencing factors or evaluation indicators in terms of your evaluation objective and intervention maturity stage.

eg. If you want to achieve objective 1 and investigate the effectiveness of the intervention, you can focus on examining patient outcomes.

Step 4: Consider the timing of evaluation and data collection in terms of evaluation requirements.

eg. If you want to investigate the intermediate changes of patients, pre- and post-test or continued monitoring during the intervention are recommended.

Step 5: Decide the study design, data collection and analysis methods regarding pre-determined evaluation objectives, timing, and indicators.

eg. If you want to collect detailed experience data with a smaller sample size, qualitative research such as interviews is recommended.

Step 6: Report your evaluation processes and results in detail and explain what theoretical or practical implications they have.

eg. If you want to inform future research, you can propose theoretical implications.

Multimedia Appendixes

The 1st & 2nd round review.
URL: http://asset.jmir.pub/assets/6b3feb5a789021c0d6a3b1864de744da.xlsx

A rapid scoping search results.
URL: http://asset.jmir.pub/assets/a8231abf153055f3548f44158bc2ca6e.xlsx

The research aims or questions of the newly eligible reviews.
URL: http://asset.jmir.pub/assets/92e0f94e08c87f1adf44948485abe7ee.xlsx

Study Characteristics and Digital Health Interventions Characteristics of Included Reviews.
URL: http://asset.jmir.pub/assets/7e4eb6a2ba5b79a22a935deba97fbc67.docx

The preferred reporting items for systematic reviews and meta-analyses methodology 2020 checklist.
URL: http://asset.jmir.pub/assets/3ef1123f1ffab22313d6690423b9b35e.docx