

Original Paper

Selection of Outcomes and Measurement Methods in Clinical Trials for Digital Therapeutics: Qualitative Analysis

Johannes Mahlmann¹, MD; Anne Mainz¹, MSc; Vera Weirauch^{1,2}, MA; Achim Mortsiefer³, Prof Dr; Sven Meister^{1,2}, Prof Dr

¹Health Care Informatics, Faculty of Health, School of Medicine, Witten/Herdecke University, Witten, Germany

²Department of Healthcare, Fraunhofer Institute for Software and Systems Engineering, Dortmund, Germany

³General Practice II and Patient-Centredness in Primary Care, Institute of General Practice and Primary Care, Faculty of Health, School of Medicine, Witten/Herdecke University, Witten, Germany

Corresponding Author:

Johannes Mahlmann, MD
Health Care Informatics, Faculty of Health
School of Medicine, Witten/Herdecke University
Pferdebachstraße 11
Witten 58448
Germany
Phone: 49 2302 92678608
Email: johannes.mahlmann@uni-wh.de

Abstract

Background: In recent years, the availability of digital therapeutic apps in medicine, digital therapeutics (DTx), has significantly increased. To evaluate the therapeutic efficacy of DTx, an increasing number of scientific studies have been conducted in this field. Several authors have pointed out that a standardized approach to DTx evaluation is relevant to improving the comparability of apps in the future. Currently, it is unclear how study authors proceed when selecting outcomes and measurement methods (O&M) for DTx studies.

Objective: The primary aim of this study was to investigate how study authors currently select O&M in DTx studies, specifically for applications delivered in the form of apps. The secondary objectives were to collect suggestions for improving the comparability of DTx apps and studies in the future and to examine whether core outcome sets (COS) could represent a possible solution for enhancing comparability.

Methods: This study was based on 19 semistructured interviews. All interview partners had prior experience in planning and conducting DTx studies. The interviews were conducted between August 2024 and November 2024. Subsequently, qualitative data analysis was performed according to Kuckartz and Rädiker using a deductive-inductive research approach.

Results: The results of this study showed that the methodological approach for selecting O&M in DTx studies was primarily based on research, as well as on personal prior experience and individual ideas. Interview participants indicated that when selecting O&M, they particularly focused on aspects such as scientific quality criteria or practical implementation and feasibility. A suggestion for achieving better comparability of DTx studies in the future was to use standardized O&M per therapeutic indication area. Almost all interviewees considered COS to be a suitable approach for improving the comparability of DTx and associated studies. However, we also identified critical perspectives that questioned their application.

Conclusions: Across professional groups, both commonalities and differences were observed in methodological approaches as well as in the selection of O&M in DTx studies. To improve the comparability of DTx interventions and trials, several approaches have been developed, for example, through the use of guidance frameworks, artificial intelligence programs, or the establishment of COS. However, further research is warranted to advance this field.

JMIR Form Res 2026;10:e86437; doi: [10.2196/86437](https://doi.org/10.2196/86437)

Keywords: digital therapeutics; DTx; digital health interventions; digital health evaluation; mobile health apps; qualitative research; core outcome set; outcome measurement; clinical efficacy trial; clinical trial methodology; measurement instruments

Introduction

Background

Digitalization has led to changes in health care at various levels [1]. Within this development, numerous terms have been established that describe the connection between digital innovations and health care. One example is the term digital therapeutics (DTx). According to the European Data Protection Supervisor [2], DTx are defined as “evidence-based therapeutic interventions driven by software to prevent, manage, or treat a medical disorder or disease. In other words, DTx are patient-facing software applications that help patients treat, prevent, or manage a disease and that have a proven clinical benefit” [2]. Thus, DTx (1) are software-based apps; (2) are intended primarily for patients rather than health care providers or third parties; (3) can be distinguished according to their primary therapeutic objective into apps intended for disease treatment, disease management, or the improvement and/or prevention of health care functions; (4) are based on concrete therapeutic concepts; and (5) have demonstrated their therapeutic efficacy through scientific studies [2-8].

In the European context, DTx are classified as medical devices and consequently assigned to risk classes I-III [9]. They are therefore subject to authorization requirements comparable to those for other medical devices or pharmaceuticals [10].

Possible technical realizations of DTx include web apps, smartphone apps, or serious games [4,11]. Currently, numerous apps are available for various medical conditions, including mental disorders, cardiovascular diseases, addiction disorders, cancer, and pain therapy [11,12]. DTx must be clearly distinguished from fitness and lifestyle apps that do not meet DTx criteria and whose efficacy has not been demonstrated in clinical studies [6,13,14].

Typically, multiple studies are carried out throughout the development and authorization of DTx. Study objectives and the selection of outcomes and measurement methods (O&M) usually differ depending on the technical development phase and scientific question. Possible research questions may include user experience, usage duration, or treatment cost reduction of an app [15-18]. As previously mentioned, one quality characteristic of DTx is that therapeutic efficacy must be demonstrated in studies [14,19]. Internationally, there is currently a lack of uniform requirements regarding what should constitute efficacy evidence and which criteria must be fulfilled for app authorization [6,9,12,20,21].

Germany has developed a structured authorization and reimbursement process for DTx through the Digital Healthcare Act and the fast-track procedure for digital health applications (DiGA). The German procedure includes requirements for data protection, user-friendliness, and demonstration of positive care effects. Study end points for efficacy demonstration may include health status, illness duration, survival, morbidity, mortality, or quality of life, with the latter being particularly patient relevant [22]. With

reference to this system, several countries have taken the German system as a model and have initiated the development of comparable frameworks [23,24]. For Europe, France with the *Prise en Charge Anticipée* model and Belgium with the National Institute for Health and Disability Insurance (INAMI-RIZIV) model are particularly noteworthy, both offering concrete requirements for authorization processes and app reimbursement [20].

The definition of a high-quality DTx depends on the evaluation criteria applied by different stakeholders, including patients, clinicians, health care payers, developers, and policymakers, whose perspectives and priorities shape their individual assessments [25]. Therefore, assessments of DTx quality vary substantially across contexts and stakeholder interests. The Digital Therapeutics Alliance (DTA), which is one of the leading international nonprofit organizations dedicated to the research and development of DTx, has emphasized that there is currently no unified, internationally accepted assessment framework that is independent of individual stakeholder interests [25]. To promote harmonization in the evaluation of DTx, the DTA published the Value Assessment and Integration Guide [25]. This guide provides a structured approach for the holistic assessment of DTx, independent of therapeutic area and evaluating stakeholder group.

A complementary approach is the DTx Real-World Evidence Framework developed by Kim et al [26], which offers a scientifically grounded and clearly structured methodology for the design, development, evaluation, and long-term monitoring of DTx. The authors demonstrate the application of this framework using a concrete example, placing particular emphasis on the design of DTx studies under real-world conditions and on key considerations in the selection of appropriate comparator groups [26].

Taken together, these approaches provide valuable guidance on how DTx can be developed and evaluated following implementation. However, these frameworks offer limited actionable guidance regarding the design of individual DTx studies and the key considerations required to enable long-term comparability of DTx based on study results.

As noted by various researchers, inconsistent or a lack of guidelines specifically addressing the selection of O&M in DTx studies can lead to methodological problems [17, 27-29]. While individual app efficacy can be demonstrated through scientific studies, inconsistent O&M across different studies severely limits comparability between studies and DTx, assuming that apps within the same treatment area should be compared [17,27-29]. Given the predicted further increase in DTx numbers and associated studies, comparability is fundamentally important for making evidence-based therapeutic decisions and answering reimbursement questions [30,31]. The use of core outcome sets (COS) has been mentioned in the literature as a potential solution for achieving a minimum level of comparability and study quality in DTx studies but has not been empirically evaluated yet [17,28].

COS are defined as “sets [that] represent the minimum that should be measured and reported in all clinical trials of a specific condition [...] The existence or use of a core outcome set does not imply that outcomes in a particular study should be restricted to those in the relevant core outcome set. Rather, there is an expectation that the core outcomes will be collected and reported, making it easier for the results of studies to be compared, contrasted and combined as appropriate; while researchers continue to explore other outcomes as well” [32].

Objectives

Although the problem of poor comparability between DTx studies and apps due to inconsistent O&M has been repeatedly noted in the literature, it remains unclear how O&M are currently selected in DTx studies. In our view, addressing these issues requires 2 key steps. First, current practices among study authors need to be examined to understand the criteria they consider when selecting O&M for DTx studies. Second, study authors with prior experience in DTx research should be asked to assess whether they consider COS to be a suitable solution or to propose alternative approaches. On the basis of this literature-derived rationale, the following 3 research objectives were developed for this study:

1. To determine how O&M are currently selected in DTx studies specifically for DTx in the form of apps
2. To collect ideas for achieving better comparability of DTx studies and apps in the future
3. To obtain an assessment of whether COS could contribute to improving comparability of DTx studies and apps

Methods

Overview

Qualitative research methods are established approaches in the context of medical research and health services research. They are particularly suitable for gaining a better understanding of individuals' decisions and thought processes. Within qualitative research, different methodological traditions and schools of thought exist. For this study, we chose the approach by Kuckartz and Rädiker as the theoretical foundation. Following this approach, qualitative content analysis was applied as the methodological procedure, as it is regarded as one of the most suitable qualitative methodologies for the research aims in the context of digital health research and was the approach with which the research team had the greatest prior experience [33]. The presentation of the study findings is guided by established qualitative research practices and recommendations and combines an objective, partly quantitative presentation with a complementary interpretive perspective on the interviewees' statements [34-37]. To ensure methodological rigor, this study follows the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Checklist 1) [38].

Sampling and Recruitment

To obtain answers to the research questions from various perspectives, we decided on purposeful sampling for recruiting interview participants until thematic saturation was reached. Inclusion criteria were participants who (1) could demonstrate topic-related research activity, (2) had published topic-related publications, (3) could demonstrate topic-related professional activity (eg, health insurance or development, distribution, or application of medical apps or DiGA), or (4) were involved in creating legal guidelines for DiGA.

These inclusion criteria were chosen in line with the overarching aim of our research project, which sought to conduct an in-depth analysis of the planning and implementation of DTx studies and to develop recommendations for improving their comparability. In our view, prior experience in this field was necessary to adequately address these research questions. Stakeholder groups, such as end users of DTx or physicians who prescribe these apps, would not have been able to address the specific research questions of this study in the same depth, as they are not directly involved in study design and implementation. We therefore excluded these groups from participation, as they were not expected to provide the insight required for this investigation. On the basis of these criteria, we contacted individuals and companies through national networks and the DiGA directory. The DiGA directory is a German BfArM (Bundesinstitut für Arzneimittel und Medizinprodukte) registry that lists manufacturers who have submitted DiGA applications, shows which applications are under review or approved, and provides patients and physicians with an overview of prescribable apps for various medical conditions [39]. When we became aware of specific DTx studies through these sources, we also contacted the study authors.

Data Collection

Problem-centered interviews according to Witzel and Reiter [40] were conducted between August 2024 and November 2024 via videoconferencing programs. At the request of participants, 2 interviews were carried out via telephone due to organizational considerations. All interviews were facilitated by a single researcher (JM) who had received formal training in qualitative interviewing.

To prepare for the interviews, we developed a semistructured interview guide (Multimedia Appendix 1). The guide's development was informed by the publications discussed in the *Introduction* section, which highlight that inconsistent or a lack of guidelines for the selection of O&M in DTx studies can lead to methodological challenges. The guide was pretested and collaboratively revised with 4 individuals (A Mainz, VW, SM, and A Mortsiefer). All contributors had several years of professional experience as well as expertise in research on and with DiGA. The pretest supported the structuring of the interview guide into clear thematic sections and helped refine it to focus on questions relevant to the research objectives.

At the beginning of each interview, the interviewer briefly introduced himself and explained the study using an example.

Interview partners were informed that this study aimed to investigate how O&M are selected in clinical studies of DTx delivered in the form of mobile apps. The subsequent interview guide could be divided into 4 content sections. The first section asked interview partners about personal characteristics (eg, age, current professional activity, and participation in DTx studies). The second interview section focused on methodological approaches and content aspects in O&M selection for DTx studies. Questions about the methodological approach aimed to understand how O&M are selected in DTx studies (eg, literature review or application oriented). Questions about content aspects addressed what they considered when selecting O&M (eg, financial aspects or clinical relevance). In the third interview section, participants were presented with statements from other authors emphasizing that, at present, meaningful comparisons across different DTx apps and across different DTx studies remain largely infeasible, as studies use heterogeneous O&M. Participants were then asked about possibilities they saw for achieving better comparability of DTx apps and studies in the future. In the fourth interview section, participants were introduced to COS as a possible solution for improving comparability and study quality based on the definition provided in the *Introduction* section and by Webbe et al [41]. Following this definition, interviewees were asked for their opinions on having COS specifically for DTx within each therapeutic area, such as depression therapy or pain therapy. After completing all interview questions, interview partners had the opportunity to express themselves on topic-related but independent matters and provide suggestions or feedback.

Data Analysis

After recording, interviews were imported into MAXQDA 2022 software (VERBI Software GmbH), automatically transcribed, and subsequently checked by 1 person (JM) for completeness, transcription accuracy, and depersonalization. Following this process, the transcripts were grouped in the MAXQDA software program according to the professional background of the interviewees to enable a comparative analysis of methodological approaches and substantive considerations in the selection of O&M across professional groups. Subsequent content processing and analysis were conducted manually in accordance with the qualitative content analysis approach of Kuckartz and Rädiker [42-44]. Methodologically, the study follows a deductive-inductive procedure that combines thematic structuring and evaluative elements, supported by thematic statements and complemented by a quantitative representation where appropriate.

The primary (deductive) coding system was developed based on the semistructured interview guide. This system was tested on 6 interviews and reevaluated by the research team (JM, VW, SM, and A Mainz). The primary coding system can be found in [Multimedia Appendix 2](#). To develop the secondary (inductive) coding system, all interviews were first coded with the primary coding system. Brief summaries were then written for all coded text segments, and subcodes were subsequently derived from these summaries. This methodological approach was guided by the recommendations of Kuckartz and Rädiker [42,44] regarding the development

of primary and secondary coding systems for qualitative research. After developing the secondary coding system, it was reevaluated by the research team (JM, SM, and VW), leading to the addition of the subcode category "Requirements for COS." The secondary coding system, including definitions and text examples, can be found in [Multimedia Appendix 3](#). To validate the completeness of the coding framework, all interview texts were reread completely by JM to identify any additional segments that could either be assigned to the existing coding system or would require the development of an additional code category. At this stage, we observed that all relevant statements made by interview participants could be assigned to existing code categories and that the categories showed sufficient conceptual distinction for the planned analysis and presentation of results. In accordance with the recommendations of Kuckartz and Rädiker [42-44], this was considered the point of thematic saturation. Although saturation was reached after 15 interviews, additional interview appointments had already been scheduled; therefore, these were conducted and included in the analysis.

For the presentation of our research findings in this publication, all interview quotations from German-speaking participants were translated into written English.

Ethical Considerations

The study procedure and the interview guide were reviewed by the Ethics Committee of Witten/Herdecke University and granted positive ethical approval on May 23, 2024 (application S-120/2024).

Prior to participation, all potential interview participants received an information package describing the aims and procedures of the study, together with a data protection information sheet and a written informed consent form. Interviews were conducted only after participants had provided written informed consent. Immediately before the interview began, participants were reminded that their participation was voluntary and that they could withdraw from the study at any time without providing a reason.

To protect participants' privacy and confidentiality, all collected data were deidentified prior to analysis. As part of the deidentification process, interviews were assigned sequential identification numbers according to the order in which they were conducted. All interview transcripts were carefully reviewed to remove or redact any information that could potentially identify the participants or reveal their place or position of employment.

Participation in this research project was entirely voluntary. Participants did not receive any financial compensation or other form of remuneration for their participation.

Results

Characterization of Interview Partners and Interviews

This qualitative research project interviewed 19 participants (men: $n=12$; women: $n=7$) based on a semistructured interview guide. For the analysis of the collected data, participants were grouped by professional role as follows: 7 researchers, 6 physicians and psychologists, and 6 entrepreneurs and consultants. Participants' ages ranged from 27 to 72 (mean 43, SD 11; range 27-72) years. On average, participants reported 17 years of professional experience and had conducted 7 DTx studies. The number of DTx studies conducted in app form ranged from 1 to 20 (median 5, IQR 4-8.5) studies. Of the 19 interview partners, 16 indicated having primarily experience conducting DTx studies in Germany. An overview of this information can be found in [Multimedia Appendix 4](#).

The average interview duration was 29 minutes, with the shortest interview lasting 10 minutes and the longest lasting 50 minutes (median 25 min).

General Results

The analysis indicated that thinking and decision-making in the planning and conduct of DTx studies were grounded in a range of methodological approaches and shaped by differing evaluations of key conceptual aspects. Reflecting both convergent and divergent approaches to the design and implementation of DTx studies, participants also articulated diverse perspectives on how DTx research might be further advanced and refined in the future.

[Figure 1](#) provides an overview of all interviews as well as the codes and subcodes identified within each interview section during the qualitative content analysis. For

completeness and ease of reference, the corresponding data are also presented in tabular format in [Multimedia Appendix 5](#).

According to Kuckartz and Rädiker [42], quantitative analyses within qualitative research projects represent a valuable means of obtaining an initial overview of qualitative data. They help identify thematic emphases as well as similarities and differences within the dataset, thereby providing the basis for the subsequent qualitative interpretation. However, no conclusions regarding the substantive significance or relative importance of the identified themes can be drawn from this quantitative overview. Such interpretations are therefore reserved for the subsequent qualitative analysis of the findings.

The figure is structured as follows. The top row lists the professional groups to which the interview participants were assigned based on their self-reported professional background. The second row contains the interview identification numbers, which were assigned according to the chronological order in which the interviews were conducted. The gray-shaded rows with larger font (eg, Methods, Requirements, or Core outcome sets) indicate the main thematic sections of the interview guide and serve to structure the figure conceptually.

The left-hand column lists the assigned codes together with their corresponding subcodes. Bold type indicates the higher-level codes within each interview section. For example, the category Methods>Measurement methods refers to the interview section in which the methodological approach to the selection of O&M is discussed, with the coded statements relating specifically to measurement methods. The left-aligned, nonbold entries listed beneath represent the respective subcodes within this coding category. The final column on the right-hand side of the figure indicates the number of interviews in which the respective subcode was assigned at least once.

Figure 1. Distribution of subcode categories across interview groups. COS: core outcome set; DTx: digital therapeutics; O&M: outcomes and measurement methods.

	Researchers							Physicians and psychologists					Entrepreneurs and consultants							
	15	13	12	11	09	04	01	03	17	16	08	06	05	18	14	10	07_2	07_1	02	Number of Interviews
Methods																				
Methods > Measurement methods																				
Own ideas and prior experience	1	1	0	0	0	0	1	1	0	1	0	1	1	0	0	0	1	1	0	9
Outcome-oriented	0	0	1	1	0	0	0	1	0	1	1	1	1	1	0	1	1	1	0	11
Research	1	1	0	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	16
Regulatory and expert guidance	1	0	0	0	0	1	0	1	0	0	0	0	1	0	1	0	1	1	1	8
Other	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Therapeutic context	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	1	0	5
Methods > Outcomes																				
Own ideas and prior experience	1	1	0	1	1	0	1	0	0	1	1	1	0	1	0	0	1	1	0	11
Reimbursement goal	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	2
Measurement method-oriented	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3
Research	0	1	1	1	0	1	1	1	1	0	1	1	0	1	1	1	1	1	1	15
Regulatory and expert guidance	1	0	1	0	0	1	1	1	0	0	0	0	0	1	1	0	0	1	0	8
Therapeutic context	1	0	1	1	1	1	0	0	0	1	1	0	1	1	1	1	1	1	0	13
Requirements																				
Requirements > Measurement methods																				
Financial aspects	1	0	1	1	1	0	0	1	1	0	0	1	1	0	0	0	0	1	0	9
Clinical relevance and comparability	1	1	0	1	1	1	1	1	0	1	1	0	0	1	1	1	1	0	1	14
Outcome	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	6
Patients and target group	0	1	0	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	6
Practical implementation and pragmatic aspects	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	1	0	1	1	15
Regulatory requirements and approval aspects	1	0	1	1	0	0	1	1	0	0	0	0	0	1	0	0	1	1	1	9
Scientific quality	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
Requirements > Outcomes																				
Financial aspects	1	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0	6
Clinical relevance and comparability	0	1	1	0	1	0	1	0	1	0	1	0	1	1	1	1	0	1	0	11
Measurement methods	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	1	1	0	0	6
Patients and target group	1	0	0	0	1	1	1	1	0	0	0	1	0	0	1	1	1	0	0	9
Practical implementation and pragmatic aspects	0	1	1	1	0	1	1	1	0	0	1	1	0	1	0	1	1	1	0	12
Regulatory requirements and approval aspects	1	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	1	1	1	7
Other	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Scientific quality	1	0	1	1	1	1	1	0	0	1	0	0	1	0	0	1	0	0	0	9
Proposals for improving comparability																				
Uniform definitions for O&M	1	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	5
Perceiving no need for change	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	4
Orientation to existing O&M resources	1	1	0	1	0	1	0	1	1	0	1	1	1	0	1	1	1	1	0	13
Regulatory requirements	1	0	0	0	1	0	0	1	1	0	0	0	0	1	1	1	0	1	0	8
Other	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2
Enhanced reporting	0	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	4
Practical study requirements	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	3
Core outcome sets																				
Core outcome sets > Advantages																				
Work relief and orientation in study planning	0	1	0	1	0	0	1	0	0	0	1	0	0	1	1	0	1	1	0	8
Better comparability	0	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	4
Minimum level of quality	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	4
Option for individual design	1	0	1	1	1	1	0	0	0	1	1	1	1	0	0	0	1	0	0	10
Security in approval process	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Core outcome sets > Disadvantages																				
Diverse study project	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	3
Missing / false security	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Inhibits innovation	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
No need for COS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	2
Additional burden	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Not universally applicable	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	2
Unequally large work amounts	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Core outcome sets > Requirements																				
Individual selection	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	3
Officially recognized	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	3
Quality guarantee	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
Scope and burden	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	3
Universally applicable	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1

Descriptive Analysis of the Code Distributions

The first section of Figure 1 relates to the interview questions addressing the methodological approach to selecting O&M in DTx studies. The interview data were analyzed separately depending on whether participants referred to the selection of Measurement methods or Outcomes.

For both the selection of Measurement methods and Outcomes, the subcode category Research was the most frequently assigned category. It was coded in 16 of the

19 interviews in relation to the selection of Measurement methods and in 15 interviews in relation to the selection of Outcomes. Thus, this category was represented in almost all interviews and across all professional groups.

In contrast, marked differences between the selection of Measurement methods and Outcomes were observed for the subcode category Therapeutic context. While this category was assigned only five times in relation to the selection of Measurement methods, it was identified in 13 interviews concerning the selection of Outcomes.

A similar pattern emerged for the subcode category Outcome-oriented, which was coded in 11 interviews addressing the selection of Measurement methods, whereas the corresponding counterpart, Measurement method-oriented, was identified in only 3 interviews concerning the selection of Outcomes. Overall, participants more frequently described the selection of measurement methods as being guided by predefined outcomes than the selection of outcomes as being determined by already established measurement methods.

Differences were also observed regarding the thematic breadth of the interviews. For example, interviews 15 and 07_01 covered a wide range of different subcode categories within this topic area, whereas interviews 02, 09, and 17 contained comparatively few distinct subcodes related to the methodological approach.

The second section of [Figure 1](#) summarizes the requirements considered by the interview participants when selecting O&M.

Regarding Measurement methods, the subcode category Scientific quality was the most frequently represented category, appearing in 18 of the 19 interviews. This was followed by Practical implementation and pragmatic aspects (15 interviews) and Clinical relevance and comparability (14 interviews).

A different distribution pattern emerged for the requirements relating to Outcomes. Here, Practical implementation and pragmatic aspects was the most frequently assigned subcode category, occurring in 12 interviews. This was followed by Clinical relevance and comparability (11 interviews) and the subcode categories Scientific quality and Patients and target group, each of which was represented in 9 interviews.

Differences between professional groups were particularly evident for the subcode category Scientific quality. Scientific quality was represented in 18 of the 19 interviews addressing the requirements for Measurement methods, indicating that this aspect was consistently considered across the dataset. In contrast, in discussions of the requirements for Outcome selection, Scientific quality was assigned predominantly in interviews with researchers, whereas it was represented considerably less frequently in interviews with physicians and psychologists as well as entrepreneurs and consultants.

Within the professional group of entrepreneurs and consultants, 2 notable patterns emerged. First, the subcode category Financial aspects was assigned only once in relation to the requirements for Measurement methods within this professional group. Second, the subcode category Regulatory requirements and approval aspects was assigned more frequently in relation to the requirements for Outcomes than in the other professional groups. Both patterns differed from the remaining professional groups, in which Regulatory requirements and approval aspects were discussed less frequently and Financial aspects more frequently.

The third section of [Figure 1](#) addresses proposals for improving the comparability of DTx studies and apps. The

most frequently represented subcode category within this interview section was Orientation to existing O&M resources. Statements assigned to this code described the use of existing O&M resources, such as guidance systems or standardized O&M approaches for specific indications, as an appropriate strategy to improve the comparability of studies and apps. This category was represented across all professional groups and in 13 of the 19 interviews.

Professional group-specific differences were particularly evident for the subcode categories Uniform definitions for O&M and Regulatory requirements. Statements referring to Uniform definitions for O&M occurred more often in interviews with researchers, whereas Regulatory requirements were more frequently assigned in interviews with entrepreneurs and consultants.

The final section of [Figure 1](#) summarizes participants' views on whether COS represent a suitable approach for improving the comparability and quality of DTx studies and apps.

Among the perceived advantages of COS, the subcode categories Option for individual design (10 interviews) and Work relief and orientation in study planning (8 interviews) were assigned most frequently. Overall, advantages were discussed more often than potential disadvantages (27 vs 11 coded segments).

The distribution of the codes also revealed distinct response patterns across interviews. In 11 of the 19 interviews, participants discussed only the advantages of COS without simultaneously mentioning potential disadvantages or requirements for their development.

Selection of O&M in DTx Studies

Overview

The subsequent sections of the results present the qualitative findings of this study. The second part of the interviews addressed the selection of O&M in DTx studies. Participants were asked about both their methodological approach to selecting O&M (ie, how they proceed when selecting O&M for a DTx study) and the content-related aspects considered during this process (ie, which criteria they consider when selecting O&M).

Research as a Foundation

Many interviewees described during the conversations that they were not aware of any universal master plan for how DTx studies should be methodologically designed. Rather, each study was understood as an individual project that required fundamental reconceptualization from the outset. Given the diversity of thematic and therapeutic approaches across different DTx apps, the initial phase of a study project was described as requiring basic orientation.

Across all professional groups, participants reported conducting a search activity as a central methodological component of this process. This activity was described as supporting the identification of O&M that might be suitable for scientifically evaluating the respective app. This practice

was also reflected in the data presented in the Methods section of [Figure 1](#), where decisions regarding outcome selection and measurement methods were most frequently grounded in prior research.

The scope and structure of the search process described by interviewees varied considerably. While some participants reported conducting rather informal and unsystematic searches, others described carrying out more structured searches in the form of systematic literature reviews in accordance with scientific quality criteria:

Well, I would specifically conduct a literature review to see which clinically validated, or clinically tested tools are available for the disease entity or application. [Interview 09, researcher]

Drawing on Established Experience

Beyond research-based approaches, interviewees reported that the selection of O&M for DTx studies was also informed by prior professional experience. The professional background of participants appeared to shape the type and relevance of experience drawn upon in this process.

Researchers predominantly referred to methodological experience, particularly regarding test procedures or study designs that they or others had previously applied in comparable research contexts:

We use proven methodologies that we have often applied and that are well known. [Interview 13, researcher]

In contrast, physicians and psychologists emphasized the integration of clinical expertise, particularly their knowledge of specific disease patterns and practical experience working with patients:

I actually often use tests from my clinical experience that have also proven to be relevant and effective in clinical studies. [Interview 16, physician and psychologist]

Similarly, entrepreneurs and consultants drew on regulatory knowledge and insights from approval processes. They also considered design strategies observed in competing DTx products with comparable therapeutic indications. In doing so, they took into account how similar apps had structured their studies, which methodological choices had been accepted in previous approval processes, and which approaches appeared to be most successful in terms of market authorization and positioning.

The substantive value of prior experience, as well as the degree of systematicity with which it was incorporated into the planning process of DTx studies, can be illustrated by the following finding. A considerable number of interviewees described developing personal databases based on accumulated experience, a pattern that was also reflected in the coded data. During the planning process, particularly when selecting

O&M, these databases enabled them to draw systematically on previously acquired knowledge and experiences to guide their decision-making:

I have gradually created a sort of database in which I collect well-known measurement instruments. Their advantages and disadvantages are listed, so that I can already make a selection to some extent. [Interview 06, physician and psychologist]

Efficiency in planning and conducting DTx studies was closely linked to both the breadth of prior experience and the degree to which this experience was systematically documented, indicating that experience functions not only as accumulated knowledge but also as an operational resource in study design.

Across groups, prior experience thus functions as a pragmatic heuristic that reduces complexity and compensates for uncertainty in study planning, albeit with different profession-specific emphases.

External Stakeholder Input and Validation

Across all professional groups, a recurring practice in the selection of O&M for DTx studies was the early exchange with external stakeholders. These included clinical experts, methodological specialists in DTx research, and regulatory authorities.

These exchanges were primarily used to discuss and validate planned study designs, particularly with regard to the appropriateness of selected end points and measurement procedures. The overarching aim was to ensure that the chosen measures were both scientifically suitable for evaluating the respective DTx app and aligned with expectations relevant for later approval and evidence generation. In this sense, external input functioned as an early orientation and validation mechanism during the study design phase.

As most interviewees had experience with the German regulatory context, particularly the BfArM framework, interactions with regulatory authorities were often described as taking place within a relatively structured approval environment. However, similar patterns were also reported in other contexts, including the United States:

So right now, with the study in the United States, we met with the FDA very often before—although they didn't really have a proper procedure yet, at least not like in Germany, but something similar. And of course, we talked a lot about it. What is the endpoint? What should the comparison group look like? Should the primary endpoint be after three or after six months? What is the minimal clinically important difference? Where does it come from? How do we use it to classify the difference between groups? Do we dichotomize it, or do we look at the group mean and that? [Interview 15, researcher]

Key Considerations Guiding the Selection of O&M in DTx Studies

Analysis of content criteria considered when selecting O&M in DTx studies revealed a differentiated picture of various priorities and requirements, partially differing between professional groups. As shown in the Requirements section of Figure 1, interviewees particularly considered aspects related to clinical relevance and comparability, practical implementation and pragmatic aspects, as well as scientific quality when selecting O&M.

Scientific Rigor

In the selection of measurement methods, scientific quality criteria emerged as a central aspect across all groups of interviewees. Participants emphasized psychometric properties, including validation of O&M for specific target populations, reliability, and indicators such as the minimal clinically important difference (MCID).

The MCID addresses a persistent challenge in medical research and patient care, namely that objectively measured parameters may show statistically significant changes during treatment, while these changes are not necessarily perceived as meaningful by patients themselves. To address this discrepancy, the development and application of MCIDs have gained increasing importance. For specific diagnostic measures or outcomes, the MCID defines the smallest change in score that is considered clinically meaningful for an individual patient [45]. In the context of planning and conducting DTx studies, MCIDs were described as particularly relevant for assessing whether a DTx app provides a meaningful therapeutic benefit for patients.

Questionnaires were reported as one of the most frequently used methods for evaluating DTx interventions. In this context, interviewees highlighted the importance of linguistic adaptation and validation across different populations. Beyond language, factors such as age-specific suitability and disease-specific relevance were described as requirements that must be covered by prior validation of the instruments before selection.

Regarding the selection of appropriate measurement instruments, interviewees highlighted the challenge of identifying tools that adequately meet the methodological and clinical requirements of the respective study:

You need to choose a sufficiently valid instrument that can measure the intended endpoints. This is often challenging, because not all instruments have all the necessary psychometric properties. [Interview 03, physician and psychologist]

Pragmatic Implementation

Beyond scientific considerations, a substantial number of statements emphasized the importance of practical constraints in the implementation of DTx study designs in real-world clinical settings. Across interviews, study planning was consistently shaped by the feasibility of integrating

methodological approaches into hospital or therapeutic environments. Available resources, such as equipment or tools, functioned as key decision criteria by enabling or constraining the selection of specific O&M:

It should be a measurement method that is simple to perform, where perhaps the necessary measuring devices are already available. [Interview 13, researcher]

A further recurring theme concerned the burden associated with study implementation for both study teams and patients. More complex study designs were described as requiring additional personnel resources or specifically trained staff, which were not always available in sufficient capacity. For patients, minimizing study-related burden was repeatedly highlighted as essential to reduce dropout and ensure continued participation throughout the study period:

So when do we assume that we can detect a change, because we actually want to measure as early and infrequently as possible to avoid burdening patients and have the lowest possible dropout. [Interview 01, researcher]

In addition, timing considerations for outcome assessment were described as relevant not only in relation to feasibility and burden but also in response to regulatory requirements. Regulatory time frames were reported to influence study planning, particularly regarding the expected time frame in which clinical improvement must be demonstrated. Consequently, measurement approaches requiring longer observation periods were considered unsuitable under such constraints.

Clinical Applicability and Relevance

Moreover, a recurring consideration in the selection of O&M was the preference for those that had already demonstrated relevance in both clinical practice and scientific research.

From a medical perspective, interviewees emphasized the importance of selecting end points that are routinely used in clinical practice. This was described as facilitating a more accurate assessment by health care professionals of whether, and to what extent, a DTx app is effective in the treatment of a specific condition and what therapeutic benefit it may provide to patients. One interviewee from the group of physicians and psychologists (interview 05) particularly highlighted that, from a clinical standpoint, hard end points—such as absolute weight reduction, improvements in joint range of motion, or prolonged survival—were considered more relevant outcome measures. In contrast, soft outcomes, such as general well-being, were described as less suitable for demonstrating the clinical effectiveness of a DTx intervention:

So that was one of the decisive criteria for me, that it is a well-studied parameter where I can really say it is also significant for patient prognosis. So not just a parameter that might change positively, but it was

important for me that this is a marker where you can really say this is significant for this patient group when it improves. [Interview 05, physician and psychologist]

From a scientific perspective, comparability across DTx and related research projects was closely linked to the use of identical or well-established O&M within a given context. Established measurement approaches were also associated with higher acceptance of study findings within the scientific community, including their likelihood of publication.

Perspectives on Improving the Comparability of DTx Studies

Overview

The focus of the third section of the interviews was on identifying approaches that could improve the comparability of DTx apps and the associated clinical studies. The qualitative analysis revealed 3 broad perspectives. First, participants proposed strategic measures aimed at establishing common methodological standards. Second, several interviewees described concrete technological solution concepts that could support greater comparability in practice. Finally, a minority of participants questioned whether additional standardization was necessary at all. An overview of the identified approaches can be derived from the corresponding subcode categories presented in the Proposals for Improving Comparability section of [Figure 1](#).

Building Comparability Through Guidance and Regulation

The primary and most frequently discussed recommendation focused on the development of guidance systems and standardized O&M tailored to specific clinical indications ([Figure 1](#)). Participants largely agreed that a pragmatic way to improve comparability would be to define a clear framework through which appropriate and unified O&M could be identified for individual research projects. This approach was described as relying on expert-driven curation to establish a baseline set of outcomes or measurement tools—potentially including 5 to 15 key markers—for specific therapeutic areas, such as depression or diabetes management.

While some respondents highlighted medical professional societies as a suitable body to organize and oversee these standards, the data suggest a broader range of possible implementation pathways. The main objective remained the establishment of broad consensus on which parameters are most relevant for a given disease, regardless of whether this consensus is led by professional societies, multistakeholder initiatives, or regulatory bodies:

So I think it's important to first agree on, at a high level, which parameters are actually important for a

given disease and should be investigated. Then, for each condition, there could be a handful— maybe five to ten or even up to fifteen—markers that play a role. [Interview 08, physician and psychologist]

But maybe professional societies should handle this, to say, okay, in the case of back pain, no matter who publishes what [no matter who conducts a DTx study], when it comes to tackling back pain, these are the required measurement tools, so that we can actually compare results. I mean, that would be the simplest approach. [Interview 06, physician and psychologist]

Recognizing that voluntary and consensus-based guidelines may not be sufficient to ensure broad implementation, a second thematic cluster emphasized the need for formal legal and regulatory requirements ([Figure 1](#)). This recommendation shifts the discussion from optional best practices toward binding requirements directly linked to market approval and reimbursement pathways.

Within this perspective, national regulatory authorities were described as important gatekeepers that could limit the range of acceptable methodological tools. Rather than leaving the selection of questionnaires entirely to study authors, regulatory bodies could define binding frameworks of approved diagnostic instruments:

It is then the task of the regulatory authorities to provide guidance so that potential comparability can be achieved. [...] they [the regulatory authorities] should say, "Okay, we accept these and those questionnaires." Everything else can be disregarded. [Interview 18, entrepreneur and consultant]

This call for regulatory intervention is further nuanced by an international dimension. Respondents highlighted that national legislative changes do not occur in isolation. Rather, they are increasingly influenced by emerging international evaluation frameworks. To prevent fragmented, conflicting national requirements that would complicate cross-border data comparison and integration, participants emphasized that future legal standardization must be harmonized internationally to ensure long-term, macroeconomic comparability (interview 03).

Architectural Archetypes for Practical Solution Concepts

Beyond these broad strategic recommendations, several participants proposed concrete technological and infrastructural solutions that could support greater comparability in practice. Although differing in complexity and scope, these concepts can be grouped into 3 architectural archetypes according to their primary mechanism, target audience, and level of systemic intervention ([Table 1](#)).

Table 1. Architectural archetypes for practical solution concepts.

Solution archetype	Primary mechanism	Target audience	Systemic intervention level
Information layer (eg, traffic light systems)	Visual reduction of multidimensional clinical data into simplified, discrete heuristic signals	Clinicians, end users, and nonexpert stakeholders	Low or downstream leaves existing trial designs untouched; standardizes output presentation.
Methodological layer (eg, knowledge repositories and GenAI ^a)	Semantic aggregation of historical trial data into active design “cookbooks” and conversational AI assistants	Study authors, R&D ^b teams, and clinical researchers	Medium or midstream standardizes researcher behavior through historical benchmarking.
Structural layer (eg, centralized platform trials)	Complete systemic institutionalization of a continuous, multiarm clinical environment using a shared control architecture	Regulatory bodies, health insurers, and all DTx ^c vendors	High or upstream eliminates decentralized trial autonomy; standardizes the environment.

^aGenAI: generative AI.^bR&D: research and development.^cDTx: digital therapeutics.

The information layer represents an approach to post hoc visual simplification. Rather than standardizing how trials are conducted, this approach aims to standardize how results are communicated. The suggested mechanism was a simplified visual heuristic, specifically a “traffic light system” (interview 01). This solution avoids the complex task of mathematically aligning heterogeneous effect sizes across different study designs. Instead, it introduces a simplified binary or ternary evaluation filter (eg, green=proven effect on the primary end point). Its main value lies in supporting nonexpert decision-makers, such as patients and prescribing physicians, by reducing clinical complexity and enabling faster comparisons of app effectiveness.

The second archetype, the methodological layer, focuses on knowledge augmentation and AI-guided support during the study design phase, aiming to reduce future study fragmentation by directing researchers toward established best practices. This category reflects a technological development that begins with static data aggregation and evolves into more dynamic, intelligent support systems. At its foundation is a knowledge repository, often described by participants as a “cookbook,” which collects data from successful historical DTx approvals. By summarizing the most relevant outcomes, measurement instruments, and target populations for specific indications, such a registry could serve as a structured reference point for future study designs and reduce arbitrary variation (interview 07_01).

Building on this data foundation, the concept can evolve from a passive repository into an active generative support system. Through the integration of conversational AI, the effort required for literature reviews and study planning could be significantly reduced. This more advanced version would allow researchers to search historical trial designs and receive synthesized recommendations for their own clinical context:

I just had the idea for something like a ChatGPT for studies, study results and study designs. So basically what we do manually, that you can type or speak somewhere and then get an overview of studies in the area of, for example, ophthalmology, eye diseases.
[Interview 07_02, entrepreneur and consultant]

The third archetype, the structural layer, proposes institutionalized platform trials as a solution targeting the broader infrastructure of clinical evidence generation. Instead of harmonizing decentralized trials after they have been completed, this concept proposes a centralized research infrastructure in the form of publicly or insurer-funded platform trials:

That’s a continuously running clinical study where new arms can be added and ended again. You need infrastructure for that. [...] insurance companies in Germany, which ultimately have to pay for the applications, conduct such a permanent trial. Whenever someone has a new app, it has to go in there. [Interview 15, researcher]

The central innovation of this archetype lies in integrating all trial variables within one permanent institutional framework. By evaluating competing DTx apps within the same study environment, each app could be assessed against the same control group, with identical outcome measures and within the same target population. In this way, the problem of comparability could be addressed directly by replacing fragmented, independent vendor-led trials with a unified research structure.

Counterperspective—Perceiving No Need for Change

Although most participants proposed measures to improve comparability, the qualitative analysis also revealed a counterperspective that questioned the need for standardization. A minority of participants explicitly defended the status quo (“Perceiving no need for change”; Figure 1), describing methodological heterogeneity not as a problem but as a natural feature of scientific practice and innovation. Drawing comparisons to the pharmaceutical sector, these respondents argued that drug approval processes also function without strict cross-trial standardization of measurement tools. From this perspective, introducing COS into the DTx field would be unnecessary and potentially constitute counterproductive regulatory overreach:

Well, I find that studies for different products are not completely comparable, I have no problem with that at all. That is scientific life, scientific practice for me. [Interview 02, entrepreneur and consultant]

Especially since I, as already outlined, rather question whether this is needed at all. This doesn't exist in the pharmaceutical sector in that sense either. [Interview 14, entrepreneur and consultant]

COS as a Promising but Complex Solution

The final section of the interviews addressed whether COS represents a suitable instrument for improving the comparability of DTx studies and apps. The statements of the interviewees revealed a highly differentiated picture. While participants generally recognized the great potential of COS to standardize and improve clinical research, they also emphasized significant practical challenges, potential risks for trial design, and a critical lack of formal regulatory backing.

The primary benefit associated with the integration of COS into the DTx ecosystem was the systematic facilitation of the study planning process. For research teams, selecting appropriate, well-established outcomes is often a complex and time-consuming task. A predefined COS serves as a practical orientation guide that reduces the initial effort required when designing a new study. By providing a clear framework of established parameters, it directly helps study teams optimize their workflow and ensures that critical clinical dimensions are not overlooked during the conceptualization phase:

If you strive to do qualitatively good research, then you will gladly accept having to meet certain criteria and might even be happy to get guidelines where you can orient yourself and also forget less. [Interview 08, physician and psychologist]

Beyond lowering the organizational workload for research teams, the widespread adoption of COS was expected to have a strong stabilizing effect on the overall quality of DTx evaluations. When individual studies consistently align with relevant core outcomes, the resulting data automatically become more homogeneous. This structural consistency directly contributes to the macro-level objective of ensuring comparability across study results from different DTx studies.

Importantly, supporters of this approach pointed out that using a COS does not mean sacrificing scientific freedom or product-specific nuances. Because a COS is conceptually defined as a minimum baseline, it explicitly allows for a flexible, 2-layered study architecture. Researchers can easily fulfill the requirements of universal comparability while maintaining the freedom to include additional, customized parameters to capture the unique features of their specific digital app:

You have a framework that ensures comparability and simultaneously freedom to include your own parameters. [Interview 16, physician and psychologist]

Despite these apparent benefits, the qualitative analysis revealed serious counternarratives that questioned the methodological consequences and the entrepreneurial viability of COS.

The first major concern targets the practical execution of blending standardized core sets with custom parameters. Critical voices caution that adding multiple product-specific outcomes on top of a standard COS is not a perfect compromise. Instead of preserving scientific freedom, this approach risks overloading the trial protocol, overcomplicating the data collection, and ultimately diluting or distorting the original focus of the research project:

But if you say you can add additional parameters, use additional measurement instruments, that dilutes the study project in my view. [Interview 14, entrepreneur and consultant]

The second, more profound challenge lies in the current gap between methodological ideal and regulatory reality. For commercial digital health entrepreneurs, investing resources into alignment with a COS is only reasonable if it mitigates business risks. However, the data highlight deep uncertainty regarding whether compliance with a community-derived COS actually guarantees acceptance by regulatory bodies. As long as national authorities (such as the BfArM) do not officially endorse or enforce these specific core outcomes, manufacturers face the risk of designing a comparable study that still fails to achieve official market clearance or reimbursement approval:

My question is rather, does it solve any problem for me as an entrepreneur, and I fear it doesn't. [...] Because that's the question again: What does the BfArM do? Because how do you want to establish the binding nature of these Core Outcomes? [...] What makes that the standard? [Interview 02, entrepreneur and consultant]

To bridge the gap between these opportunities and risks, the interviewees formulated concrete, nonnegotiable requirements that must be met for COS to achieve routine implementation in the research landscape.

The first requirement relates to fairness across indications. Participants stressed that the structural scope of COS must remain equitable and balanced across different therapeutic areas. If different medical fields develop core sets with heavily contrasting workloads—where manufacturers in one clinical indication only have to collect 3 outcomes, while competitors in another field are forced to track 5 or more—it would create market distortions. To achieve broad acceptance among developers, the procedural and financial burden of data collection must not result in an unfair disadvantage for specific treatment areas:

At the latest, manufacturers would say, how can it be that some [...] let's say only have to collect three points for their Core Set and others have to collect five points or so. So I imagine that as a challenge. [Interview 03, physician and psychologist]

The second requirement relates to pragmatic feasibility. Operational brevity is identified as a vital success factor. Because a COS is often envisioned as an additive framework rather than a total replacement for product-specific metrics, it must be kept as short as possible. In fast-paced digital health environments, the feasibility of assessment procedures depends on keeping the burden on both patients and study teams as low as possible. If a COS substantially increases the frequency or duration of assessments, its implementation may become impractical in routine clinical practice:

And if it's feasible, I mean Core Outcome Sets have to be short, because if everyone should do this and partly additionally, then it can't multiply assessment times by five. [Interview 15, researcher]

Discussion

Principal Findings

Previous scientific literature has primarily focused on the effectiveness of individual DTx [46-49], cost-benefit or health economic evaluations [50-53], and factors facilitating or hindering their adoption and integration into clinical care [54-56]. As outlined in the *Introduction* section, previous studies have likewise emphasized the considerable heterogeneity in O&M used across DTx effectiveness studies [17,27-29]. However, little was known about how DTx studies are conceptually designed in practice and which methodological and content-related considerations guide this process. This knowledge gap was addressed in the present research. The results of this study provided an objectification of methodological approaches and content aspects of how O&M were selected in DTx studies, as well as a collection of suggestions for achieving better comparability of DTx studies and apps in the future.

In summary of the results addressing our first research question, we can conclude that the methodological approach for selecting O&M in DTx studies across all professional groups was primarily guided by research activity. Conducting an initial literature review during the early stages of study planning reflects established practice and aligns with the standards for developing high-quality clinical studies, even outside the DTx context [15,57-59]. However, when interpreting these findings, it should be noted that one inclusion criterion for participation in this study was prior DTx research activity. This criterion may have contributed to the prominence of research as a theme in the results.

The analysis of subcategories regarding the content aspects considered when selecting O&M revealed both commonalities and differing emphases. Overall, an imbalance in the total number of coded statements was observed. Within the group

of interviewees classified as researchers, substantially more statements were coded compared with the other professional groups. It should also be noted that the professional groups differed in size. The physicians and psychologists group and the entrepreneurs and consultants group were based on 6 interviews each, whereas the researchers group was based on 7 interviews. On the basis of the present data, slight tendencies can be observed regarding which substantive aspects different professional groups emphasize when selecting O&M in DTx studies. However, these tendencies are not pronounced enough to support the conclusion that systematic, profession-related differences exist in the content aspects considered when selecting O&M.

In comparison with findings reported by other authors, the statements of our interview partners showed clear convergence. Ataiy et al [60] investigated the challenges faced by DiGA manufacturers when seeking approval of DTx within the German fast-track process. Their study participants identified the demonstration of a positive health care effect through scientific studies as the major challenge, with particular emphasis on the implementation of the study protocol, the recruitment of study participants, and the coordination of logistics. Similar challenges related to the practical implementation of planned study designs were also reported by our participants and were coded under the subcode Practical implementation and pragmatic aspects in the selection of O&M.

Furthermore, the statements of the interview partners aligned with the DTA's emphasis on the use of clinically relevant and previously published end points in DTx studies [31]. However, contrary to the DTA's recommendation to select patient-relevant end points whenever possible, such considerations were reported only rarely across the interview groups [31]. Especially within the physician and psychologist group, only 2 interviewees indicated that they consider aspects related to patients and target groups when selecting outcomes. Compared to other code categories, these statements were mentioned substantially less frequently. The limited incorporation of patient and end user perspectives into the development and evaluation of DTx has repeatedly been identified as a critical limitation in the existing literature [61, 62]. Insufficient attention to co-design aspects can substantially limit the uptake and long-term use of DTx in real-world settings. A lack of alignment with users' needs and everyday contexts may undermine sustained engagement [61,63-65]. Possible approaches for involving end users in the development and evaluation of DTx include focus group interviews, think-aloud strategies, surveys, and persona building, in which a persona represents a key end user [66].

The third section of the interviews aimed to identify suitable solution approaches for improving the comparability of DTx studies and apps. As pointed out by McLeod et al [67], the selection of appropriate end points for clinical trials of health care interventions has become increasingly challenging due to the growing diversity of study designs and the differing interests of end user groups. In their view, there is a lack of guidance for selecting suitable end points in clinical trials. Similar to these views, the most frequently

expressed response by our interview participants was the need to develop guidance systems for O&M selection in DTx studies to improve comparability. This appears well suited as it is independent of country-specific legal requirements and enables international cooperation.

A similar approach is reflected in the interviewees' suggestion to develop a registry or AI system to support the selection of O&M for DTx studies. While stricter legal frameworks and regulations must be understood as static concepts, a registry documenting which types of O&M have been used in other DTx studies, or an AI system instead provides a space and opportunity for exchanging ideas and approaches to the design of DTx studies among study investigators. In particular, the concept of an interactive AI reflects the principles of agile software research. The development and evaluation of a DTx software app cannot be divided into independent phases. Rather, they mutually influence each other and form a cyclical process [18,26]. An interactive AI generates a response to a user query while simultaneously allowing the user to engage with that response by suggesting modifications or corrections. From this perspective, both proposals represent feasible and practice-oriented solution concepts. To our knowledge, there is currently no open-access registry or AI system specifically designed for DTx studies that provides study information on investigated medical conditions and applied outcomes and measurements.

Nevertheless, the participants' calls for more uniform international standards for development and approval, as well as their demands for more comprehensive and detailed reporting of DTx studies, were consistent with the statements of other authors [29,68,69]. For example, Seo et al [29] emphasized the need for standardized research designs and consistent methodological approaches to minimize confounding factors and enable clearer conclusions regarding the effectiveness of DTx across different disease areas. Similarly, Yao et al [68] highlighted that standardized and comprehensive reporting improves the reliability of study results and helps regulators evaluate interventions while protecting patient safety and rights. As outlined in the *Introduction* section, a solution frequently proposed in the literature to achieve these goals is the use of COS specifically designed for DTx studies.

According to the assessments of our interview participants, a major advantage of COS would lie in their potential to simplify research workflows and accelerate study planning. Previous work by Clayton et al [58] showed that the considerable time investment required often discourages researchers from conducting more comprehensive literature searches during study planning. As outlined in the literature and anticipated by our interview participants, COS could serve as a valuable tool to both ensure a minimum standard of study quality and support researchers in the planning process, including the identification of appropriate O&M [70,71]. Furthermore, the use of COS enables more efficient aggregation of study results in subsequent meta-analyses [41,71,72].

However, the results of our study provide a critically evaluative perspective on the use of COS in DTx studies, which has not yet been described in the literature. Our interview participants emphasized that COS should maintain an appropriate scope to allow their additional integration into study designs without placing an excessive burden on researchers or participants. They also highlighted the importance of regulatory acceptance and noted that COS need to be recognized by the authorities responsible for DTx approval. If such acceptance cannot be guaranteed, COS may still offer a theoretically valuable tool for improving comparability and quality across DTx studies, but they would not resolve an important challenge faced by study authors, namely the ongoing uncertainty about whether their DTx approval studies will be accepted by regulatory bodies. When interpreting these statements, it should be considered that they primarily reflect the perspective of study leads who have a direct interest in the acceptance and approval of their own DTx apps and study results. To obtain a more comprehensive understanding, future research should therefore examine the use of COS from the perspective of regulatory authorities to incorporate and assess their views as well.

To date, to the best of our knowledge, no COS have been specifically developed for DTx. An unanswered question is who takes on or coordinates this task. As noted by 1 interview participant, it can be hypothetically assumed that content interests regarding which O&M are considered relevant, and the level of detail required for corresponding COS, vary by stakeholder group. To avoid overrepresentation of individual interests, actors from all disciplines, including medical societies, researchers, insurance companies, politicians, providers, and patient representatives, should be involved in the development of COS [67,73,74].

Limitations

Several limitations should be considered when interpreting the findings of this study. First, the results are based on a small cohort of interviewees, most of whom had experience conducting DTx studies within the DiGA authorization context in Germany. While this focus provided valuable insights due to Germany's highly structured DTx authorization process, perspectives from study authors outside Germany were underrepresented.

Second, prior experience in conducting DTx studies was an inclusion criterion for participation. This may have introduced selection bias, as individuals actively engaged in research may be overrepresented. In addition, not all professional groups involved in DTx development, authorization, and financing could be included.

Conclusions

The scientific contribution of this study lies in its detailed examination of the methodological approaches underlying the design of DTx studies, as well as the key conceptual aspects considered during their development. According to our findings, the methodological approach to selecting O&M in studies designed to demonstrate the clinical efficacy of DTx in the form of apps was likewise primarily based on research,

as well as on personal prior experience and individual ideas. When selecting O&M, study authors particularly emphasized scientific quality criteria, alongside considerations of practical implementation and feasibility.

In addition, we explored the hypothesis frequently discussed in the literature that the use of COS may reduce heterogeneity in outcomes and measurement instruments and improve the comparability of DTx studies and apps. By directly interviewing authors of DTx studies, we assessed whether COS are perceived as a feasible and appropriate solution from the perspective of those actively involved in DTx research. Further research is required to clarify how COS for DTx can be developed, disseminated, and officially recognized depending on the therapeutic indication area.

Beyond evaluating COS as a proposed solution, our findings also point to additional potential approaches for

addressing heterogeneity, including the use of AI-assisted systems, standardized outcome measures and instruments, and other methodological guidance frameworks.

Future DTx development requires a balance between research freedom, provider interests, and clear, uniform international authorization standards to ensure therapeutic quality and comparability [55,75]. Excessive regulatory hurdles may hinder innovation, causing research and development to shift to countries with lower requirements or be discontinued, while overly low standards risk authorizing DTx without sufficient efficacy or relevant O&M, misleading users and generating unnecessary costs for health insurance companies and end users. COS development does not end with publication. Ongoing reevaluation and adaptation are needed to maintain quality and reflect technological and medical progress [74].

Acknowledgments

AI tools were used solely for language editing and translation after completion of the original German manuscript, specifically ChatGPT (OpenAI) and DeepL (DeepL SE). The authors reviewed and edited all AI-assisted content and take full responsibility for the final version of the manuscript. No AI-based tools were used in the study design, development of the interview guide, data collection, data analysis, or interpretation of the findings.

Funding

This study did not receive any external funding from public, commercial, or nonprofit sources.

Data Availability

Due to data protection and ethical considerations, the datasets generated and analyzed during this study are not publicly available. Reasonable requests for access may be directed to the corresponding author.

Authors' Contributions

JM contributed to the study conception, data collection, data analysis, and drafting of the manuscript. VW and A Mainz contributed to the development of the interview guide and the qualitative coding frameworks for data analysis. SM contributed to the study conception and drafting of the manuscript and provided overall supervision. A Mortsiefer critically reviewed the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Semistructured interview guide.

[\[DOCX File \(Microsoft Word File\), 38 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Primary coding system (details of the primary coding framework used in the qualitative analysis).

[\[DOCX File \(Microsoft Word File\), 15 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Secondary coding system—comprehensive secondary coding system with definitions and text examples.

[\[XLSX File \(Microsoft Excel File\), 30 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Participant characterization—table showing demographic and professional details of the interview participants.

[\[DOCX File \(Microsoft Word File\), 17 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Distribution of subcode categories across interview groups.

[\[XLSX File \(Microsoft Excel File\), 23 KB-Multimedia Appendix 5\]](#)

Checklist 1

COREQ checklist.

[\[XLSX File \(Microsoft Excel File\), 13 KB-Checklist 1\]](#)**References**

1. Stoumpos AI, Kitsios F, Talias MA. Digital transformation in healthcare: technology acceptance and its applications. *Int J Environ Res Public Health*. Feb 15, 2023;20(4):3407. [doi: [10.3390/ijerph20043407](https://doi.org/10.3390/ijerph20043407)] [Medline: [36834105](#)]
2. Digital therapeutics (DTx). European Data Protection Supervisor. 2025. URL: <https://www.edps.europa.eu/press-publications/publications/techsonar/digital-therapeutics-dtx> [Accessed 2025-08-24]
3. What is a DTx? Digital Therapeutics Alliance. 2025. URL: <https://dtxalliance.org/understanding-dtx/what-is-a-dtx/> [Accessed 2025-08-24]
4. Fürstenau D, Gersch M, Schreiter S. Digital therapeutics (DTx). *Bus Inf Syst Eng*. Jun 2023;65(3):349-360. [doi: [10.1007/s12599-023-00804-z](https://doi.org/10.1007/s12599-023-00804-z)]
5. ISO/TR 11147:2023(en) health informatics — personalized digital health — digital therapeutics health software systems. International Organization for Standardization. URL: <https://www.iso.org/obp/ui/#iso:std:iso:tr:11147:ed-1:v1:en> [Accessed 2025-09-01]
6. Wang C, Lee C, Shin H. Digital therapeutics from bench to bedside. *NPJ Digit Med*. Mar 10, 2023;6(1):38. [doi: [10.1038/s41746-023-00777-z](https://doi.org/10.1038/s41746-023-00777-z)] [Medline: [36899073](#)]
7. Guidance on the review and approval of digital therapeutics (DTx) (revision). Ministry of Food and Drug Safety(Korea). 2020. URL: https://www.mfds.go.kr/eng/brd/m_40/view.do?seq=72624&srchFr=&srchTo=&srchWord=&srchTp=&itm_seq_1=0&itm_seq_2=0&multi_itm_seq=0&company_cd=&company_nm=&page=1 [Accessed 2025-10-09]
8. DTx product categories. Digital Therapeutics Alliance. 2021. URL: https://dtxalliance.org/wp-content/uploads/2021/01/DTA_FS_DTx-Product-Categories_010521.pdf [Accessed 2026-02-18]
9. Arcà E, Heldt D, Smith M. Comparison of health technology assessments for digital therapeutics in Germany, the United Kingdom and France. *Digit Health*. 2025;11:20552076241308704. [doi: [10.1177/20552076241308704](https://doi.org/10.1177/20552076241308704)] [Medline: [39845519](#)]
10. Chung JY. Digital therapeutics and clinical pharmacology. *Transl Clin Pharmacol*. Mar 2019;27(1):6-11. [doi: [10.12793/tcp.2019.27.1.6](https://doi.org/10.12793/tcp.2019.27.1.6)] [Medline: [32055575](#)]
11. Phan P, Mitragotri S, Zhao Z. Digital therapeutics in the clinic. *Bioeng Transl Med*. Jul 2023;8(4):e10536. [doi: [10.1002/btm2.10536](https://doi.org/10.1002/btm2.10536)] [Medline: [37476062](#)]
12. Liang J, Fang Q, Jiao X, et al. Approved trends and product characteristics of digital therapeutics in four countries. *NPJ Digit Med*. May 26, 2025;8(1):308. [doi: [10.1038/s41746-025-01660-9](https://doi.org/10.1038/s41746-025-01660-9)] [Medline: [40419648](#)]
13. Loughheed T. How “digital therapeutics” differ from traditional health and wellness apps. *CMAJ*. Oct 28, 2019;191(43):E1200-E1201. [doi: [10.1503/cmaj.1095801](https://doi.org/10.1503/cmaj.1095801)] [Medline: [31659065](#)]
14. Wade B, Abraham J, Coder M, et al. Classification of digital health technologies. Digital Therapeutics Alliance; 2023. URL: <https://dtxalliance.org/wp-content/uploads/2023/06/Guidance-to-Industry-Classification-of-Digital-Health-Technologies-2023Jun05.pdf> [Accessed 2025-09-04]
15. Agarwal P, Gordon D, Griffith J, et al. Assessing the quality of mobile applications in chronic disease management: a scoping review. *NPJ Digit Med*. Mar 10, 2021;4(1):46. [doi: [10.1038/s41746-021-00410-x](https://doi.org/10.1038/s41746-021-00410-x)] [Medline: [33692488](#)]
16. Kernebeck S, Scheibe M, Sinha M, et al. Digitale Gesundheitsinterventionen entwickeln, evaluieren und implementieren (Teil I) – Diskussionspapier der Arbeitsgruppe Digital Health des Deutschen Netzwerk Versorgungsforschung (DNVF). *Gesundheitswesen*. Jan 2023;85(1):58-64. [doi: [10.1055/a-1933-2779](https://doi.org/10.1055/a-1933-2779)] [Medline: [36446615](#)]
17. Kernebeck S, Scheibe M, Sinha M, et al. Erratum: Digitale Gesundheitsinterventionen entwickeln, evaluieren und implementieren (Teil I) – Diskussionspapier der Arbeitsgruppe Digital Health des Deutschen Netzwerk Versorgungsforschung (DNVF). *Gesundheitswesen*. Jan 2023;85(1):e1-e1. [doi: [10.1055/a-1996-6141](https://doi.org/10.1055/a-1996-6141)] [Medline: [36446616](#)]
18. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for digital health interventions. *it - Information Technology*. Oct 25, 2019;61(5-6):253-263. [doi: [10.1515/itit-2019-0019](https://doi.org/10.1515/itit-2019-0019)]
19. Huh KY, Oh J, Lee S, Yu KS. Clinical evaluation of digital therapeutics: present and future. *Healthc Inform Res*. Jul 2022;28(3):188-197. [doi: [10.4258/hir.2022.28.3.188](https://doi.org/10.4258/hir.2022.28.3.188)] [Medline: [35982593](#)]
20. Fassbender A, Donde S, Silva M, et al. Adoption of digital therapeutics in Europe. *Ther Clin Risk Manag*. 2024;20:939-954. [doi: [10.2147/TCRM.S489873](https://doi.org/10.2147/TCRM.S489873)] [Medline: [39741688](#)]
21. Shin HJ, Cho IT, Choi WS, Kim HR, Kang MB, Yang WJ. Digital therapeutics in Korea: current status, challenges, and future directions - a narrative review. *J Yeungnam Med Sci*. 2025;42(8):8. [doi: [10.12701/jyms.2024.01004](https://doi.org/10.12701/jyms.2024.01004)] [Medline: [39551075](#)]

22. Das fast-track-verfahren für digitale gesundheitsanwendungen (diga) nach § 139e SGB v ein leitfaden für hersteller, leistungserbringer und anwender. Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM); 2020.
23. Mäder M, Timpel P, Schönfelder T, et al. Evidence requirements of permanently listed digital health applications (DiGA) and their implementation in the German DiGA directory: an analysis. *BMC Health Serv Res*. Apr 17, 2023;23(1):369. [doi: [10.1186/s12913-023-09287-w](https://doi.org/10.1186/s12913-023-09287-w)] [Medline: [37069592](https://pubmed.ncbi.nlm.nih.gov/37069592/)]
24. Schramm L, Carbon CC. Critical success factors for creating sustainable digital health applications: a systematic review of the German case. *Digit Health*. 2024;10:20552076241249604. [doi: [10.1177/20552076241249604](https://doi.org/10.1177/20552076241249604)] [Medline: [38665883](https://pubmed.ncbi.nlm.nih.gov/38665883/)]
25. DTx value assessment & integration guide version 2.0. Digital Therapeutics Alliance; 2022. URL: https://dtxalliance.org/wp-content/uploads/2022/05/DTx-Value-Assessment-Guide_May-2022.pdf [Accessed 2026-02-26]
26. Kim M, Patrick K, Nebeker C, et al. The digital therapeutics real-world evidence framework: an approach for guiding evidence-based digital therapeutics design, development, testing, and monitoring (preprint). *J Med Internet Res*. Preprint posted online on 2023. [doi: [10.2196/preprints.49208](https://doi.org/10.2196/preprints.49208)]
27. Jiang N, Yu X, Yang Y, et al. Digital therapeutics in China: comprehensive review. *J Med Internet Res*. May 27, 2025;27:e70955. [doi: [10.2196/70955](https://doi.org/10.2196/70955)] [Medline: [40424619](https://pubmed.ncbi.nlm.nih.gov/40424619/)]
28. Sippli K, Deckert S, Schmitt J, Scheibe M. Healthcare effects and evidence robustness of reimbursable digital health applications in Germany: a systematic review. *NPJ Digit Med*. Aug 1, 2025;8(1):495. [doi: [10.1038/s41746-025-01879-6](https://doi.org/10.1038/s41746-025-01879-6)] [Medline: [40750660](https://pubmed.ncbi.nlm.nih.gov/40750660/)]
29. Seo YC, Yong SY, Choi WW, Kim SH. Meta-analysis of studies on the effects of digital therapeutics. *J Pers Med*. Jan 30, 2024;14(2):157. [doi: [10.3390/jpm14020157](https://doi.org/10.3390/jpm14020157)] [Medline: [38392592](https://pubmed.ncbi.nlm.nih.gov/38392592/)]
30. Degun R, Madaan P, Ahmad S. The time is now for digital therapeutics. *EY Parthenon*; 2025. URL: <https://www.ey.com/content/dam/ey-unified-site/ey-com/en-uk/industries/life-sciences/documents/ey-the-time-is-now-for-digital-therapeutics.pdf> [Accessed 2025-10-13]
31. Setting the stage for a fit-for-purpose DTx evidentiary standard. Digital Therapeutics Alliance. 2022. URL: https://dtxalliance.org/wp-content/uploads/2022/12/DTA-Clinical-Evidence-Paper_12.22.pdf [Accessed 2025-09-08]
32. About COMET. COMET Initiative. 2025. URL: <https://www.comet-initiative.org/About> [Accessed 2025-09-15]
33. Harrison Ginsberg K, Babbott K, Serlachius A. Exploring participants' experiences of digital health interventions with qualitative methods: guidance for researchers. *J Med Internet Res*. Nov 28, 2024;26:e62761. [doi: [10.2196/62761](https://doi.org/10.2196/62761)] [Medline: [39607999](https://pubmed.ncbi.nlm.nih.gov/39607999/)]
34. Mishra S, Dey AK. Understanding and identifying 'themes' in qualitative case study research. *South Asian Journal of Business and Management Cases*. Dec 2022;11(3):187-192. [doi: [10.1177/22779779221134659](https://doi.org/10.1177/22779779221134659)]
35. Liu P, Li X, Zhang XM. Healthcare professionals' and patients' assessments of listed mobile health apps in China: a qualitative study. *Front Public Health*. 2023;11:1220160. [doi: [10.3389/fpubh.2023.1220160](https://doi.org/10.3389/fpubh.2023.1220160)] [Medline: [37780445](https://pubmed.ncbi.nlm.nih.gov/37780445/)]
36. Schroeder T, Haug M, Georgiou A, Seaman K, Gewald H. Evidence of how physicians and their patients adopt mHealth apps in Germany: exploratory qualitative study. *JMIR Mhealth Uhealth*. Jan 17, 2024;12:e48345. [doi: [10.2196/48345](https://doi.org/10.2196/48345)] [Medline: [38231550](https://pubmed.ncbi.nlm.nih.gov/38231550/)]
37. Neunaber T, Mortsiefer A, Meister S. Dimensions and subcategories of digital maturity in general practice: qualitative study. *J Med Internet Res*. Dec 19, 2024;26:e57786. [doi: [10.2196/57786](https://doi.org/10.2196/57786)] [Medline: [39699948](https://pubmed.ncbi.nlm.nih.gov/39699948/)]
38. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Oxford Academic*. URL: <https://academic.oup.com/view-large/27217733> [Accessed 2025-10-03]
39. Finden Sie die passende digitale Gesundheitsanwendung. Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). 2025. URL: <https://diga.bfarm.de/de> [Accessed 2025-09-15]
40. Witzel A, Reiter H. *The Problem-Centred Interview: Principles and Practice*. SAGE Publications Ltd; 2012. [doi: [10.4135/9781446288030](https://doi.org/10.4135/9781446288030)] ISBN: 978-1-4462-8803-0
41. Webbe J, Sinha I, Gale C. Core outcome sets. *Arch Dis Child Educ Pract Ed*. Jun 2018;103(3):163-166. [doi: [10.1136/archdischild-2016-312117](https://doi.org/10.1136/archdischild-2016-312117)] [Medline: [28667046](https://pubmed.ncbi.nlm.nih.gov/28667046/)]
42. Kuckartz U, Rädiker S. *Qualitative Content Analysis: Methods, Practice and Software*. SAGE Publications Ltd; 2023. [doi: [10.4135/9781036212940](https://doi.org/10.4135/9781036212940)] ISBN: 978-1-0362-1294-0
43. Kuckartz U, Rädiker S. *Fokussierte Interviewanalyse mit MAXQDA: Schritt für Schritt 2 Aufl.* Springer Fachmedien; 2024. [doi: [10.1007/978-3-658-40212-9](https://doi.org/10.1007/978-3-658-40212-9)] ISBN: 978-3-658-40211-2
44. Kuckartz U, Rädiker S. *Qualitative inhaltsanalyse*. In: *Methoden, Praxis, Computerunterstützung*. Beltz Juventa; 2022:978-3. URL: <https://content-select.com/de/portal/media/view/5e623532-20b8-4f33-b19e-4a1db0dd2d03> [Accessed 2024-08-19] ISBN: 978-3-7799-5533-7
45. Rennard SI. Minimal clinically important difference, clinical perspective: an opinion. *COPD: Journal of Chronic Obstructive Pulmonary Disease*. Jan 2005;2(1):51-55. [doi: [10.1081/COPD-200050641](https://doi.org/10.1081/COPD-200050641)]

46. Meyer B, Betz LT, Jacob GA, et al. Effectiveness of a digital lifestyle management intervention (levidex) to improve quality of life in people with multiple sclerosis: results of a randomized controlled trial. *BMC Neurol*. Sep 16, 2024;24(1):347. [doi: [10.1186/s12883-024-03843-5](https://doi.org/10.1186/s12883-024-03843-5)] [Medline: [39285343](https://pubmed.ncbi.nlm.nih.gov/39285343/)]
47. Twomey C, O'Reilly G, Bültmann O, Meyer B. Effectiveness of a tailored, integrative internet intervention (deprexis) for depression: updated meta-analysis. *PLoS One*. 2020;15(1):e0228100. [doi: [10.1371/journal.pone.0228100](https://doi.org/10.1371/journal.pone.0228100)] [Medline: [31999743](https://pubmed.ncbi.nlm.nih.gov/31999743/)]
48. Xu J, Ni H, Zhan H, et al. Efficacy of digital therapeutics for perioperative management in patients with lung cancer: a randomized controlled trial. *BMC Med*. Mar 28, 2025;23(1):186. [doi: [10.1186/s12916-025-04012-2](https://doi.org/10.1186/s12916-025-04012-2)] [Medline: [40155969](https://pubmed.ncbi.nlm.nih.gov/40155969/)]
49. Park SY, On SW, Yi S, et al. Evaluating the efficacy of a digital therapeutic intervention for temporomandibular disorders: multicenter, randomized, sham-controlled trial. *J Med Internet Res*. Oct 24, 2025;27:e83545. [doi: [10.2196/83545](https://doi.org/10.2196/83545)] [Medline: [41135079](https://pubmed.ncbi.nlm.nih.gov/41135079/)]
50. Lewkowicz D, Wohlbrandt AM, Bottinger E. Digital therapeutic care apps with decision-support interventions for people with low back pain in Germany: cost-effectiveness analysis. *JMIR Mhealth Uhealth*. Feb 7, 2022;10(2):e35042. [doi: [10.2196/35042](https://doi.org/10.2196/35042)] [Medline: [35129454](https://pubmed.ncbi.nlm.nih.gov/35129454/)]
51. Liu T, Zhan Y, Chen S, Zhang W, Jia J. Cost-effectiveness analysis of digital therapeutics for home-based cardiac rehabilitation for patients with chronic heart failure: model development and data analysis. *Cost Eff Resour Alloc*. Nov 6, 2023;21(1):82. [doi: [10.1186/s12962-023-00489-x](https://doi.org/10.1186/s12962-023-00489-x)] [Medline: [37932748](https://pubmed.ncbi.nlm.nih.gov/37932748/)]
52. Park H, Jang M, Kim D, et al. Cost-utility analysis and value-based pricing of digital therapeutics for pulmonary rehabilitation in chronic respiratory disease: economic evaluation based on a randomized controlled trial. *J Med Internet Res*. Dec 15, 2025;27:e73739. [doi: [10.2196/73739](https://doi.org/10.2196/73739)] [Medline: [41397244](https://pubmed.ncbi.nlm.nih.gov/41397244/)]
53. Sapanel Y, Tadeo X, Brenna CTA, et al. Economic evaluation associated with clinical-grade mobile app-based digital therapeutic interventions: systematic review. *J Med Internet Res*. Aug 1, 2023;25:e47094. [doi: [10.2196/47094](https://doi.org/10.2196/47094)] [Medline: [37526973](https://pubmed.ncbi.nlm.nih.gov/37526973/)]
54. Panitz K, Wagenaar C, Sahan F, et al. Perspectives on digital therapeutic prescribing: a qualitative study among German psychological psychotherapists. *Front Digit Health*. 2026;8:1656614. [doi: [10.3389/fdgth.2026.1656614](https://doi.org/10.3389/fdgth.2026.1656614)] [Medline: [41737814](https://pubmed.ncbi.nlm.nih.gov/41737814/)]
55. Prodan A, Deimel L, Ahlqvist J, et al. Success factors for scaling up the adoption of digital therapeutics towards the realization of P5 medicine. *Front Med (Lausanne)*. 2022;9:854665. [doi: [10.3389/fmed.2022.854665](https://doi.org/10.3389/fmed.2022.854665)] [Medline: [35492346](https://pubmed.ncbi.nlm.nih.gov/35492346/)]
56. Kendziorra J, Seerig KH, Winkler TJ, Gewald H. From awareness to integration: a qualitative interview study on the impact of digital therapeutics on physicians' practices in Germany. *BMC Health Serv Res*. Apr 18, 2025;25(1):568. [doi: [10.1186/s12913-025-12656-2](https://doi.org/10.1186/s12913-025-12656-2)] [Medline: [40247268](https://pubmed.ncbi.nlm.nih.gov/40247268/)]
57. An MW, Duong Q, Le-Rademacher J, Mandrekar SJ. Principles of good clinical trial design. *J Thorac Oncol*. Aug 2020;15(8):1277-1280. [doi: [10.1016/j.jtho.2020.05.005](https://doi.org/10.1016/j.jtho.2020.05.005)] [Medline: [32417343](https://pubmed.ncbi.nlm.nih.gov/32417343/)]
58. Clayton GL, Smith IL, Higgins JPT, et al. The INVEST project: investigating the use of evidence synthesis in the design and analysis of clinical trials. *Trials*. May 15, 2017;18(1):219. [doi: [10.1186/s13063-017-1955-y](https://doi.org/10.1186/s13063-017-1955-y)] [Medline: [28506284](https://pubmed.ncbi.nlm.nih.gov/28506284/)]
59. Sandercock P, Whiteley W. How to do high-quality clinical research 1: first steps. *Int J Stroke*. Feb 2018;13(2):121-128. [doi: [10.1177/1747493017750923](https://doi.org/10.1177/1747493017750923)] [Medline: [29299959](https://pubmed.ncbi.nlm.nih.gov/29299959/)]
60. Ataiy O, Burniki M, Sohrabi K, Groß V, Scholtes M, Schudt F. Digital health applications in Germany: explorative pilot study on the challenges of approval from the perspective of manufacturers. *Digit Health*. 2024;10:20552076241302676. [doi: [10.1177/20552076241302676](https://doi.org/10.1177/20552076241302676)] [Medline: [39711743](https://pubmed.ncbi.nlm.nih.gov/39711743/)]
61. Mimoso I, Figueiredo T, Midão L, et al. Co-creation in the development of digital therapeutics: a narrative review. *Int J Environ Res Public Health*. Nov 28, 2024;21(12):1589. [doi: [10.3390/ijerph21121589](https://doi.org/10.3390/ijerph21121589)] [Medline: [39767430](https://pubmed.ncbi.nlm.nih.gov/39767430/)]
62. Baines R, Bradwell H, Edwards K, et al. Meaningful patient and public involvement in digital health innovation, implementation and evaluation: a systematic review. *Health Expect*. Aug 2022;25(4):1232-1245. [doi: [10.1111/hex.13506](https://doi.org/10.1111/hex.13506)] [Medline: [35526274](https://pubmed.ncbi.nlm.nih.gov/35526274/)]
63. Weirauch V, Soehnchen C, Burmann A, Meister S. Methods, indicators, and end-user involvement in the evaluation of digital health interventions for the public: scoping review. *J Med Internet Res*. May 31, 2024;26(1):e55714. [doi: [10.2196/55714](https://doi.org/10.2196/55714)] [Medline: [38819891](https://pubmed.ncbi.nlm.nih.gov/38819891/)]
64. Hochmuth A, Exner AK, Dockweiler C. Implementierung und partizipative Gestaltung digitaler Gesundheitsinterventionen. *Bundesgesundheitsbl*. Feb 2020;63(2):145-152. [doi: [10.1007/s00103-019-03079-6](https://doi.org/10.1007/s00103-019-03079-6)]
65. de la Torre E, Montane C, Rubio O, Sampietro-Colom L, Camacho-Mahamud A, Grau-Corral I. Integrating patient perspectives into the digital health technology readiness framework: Delphi study. *J Med Internet Res*. Aug 20, 2025;27:e71600. [doi: [10.2196/71600](https://doi.org/10.2196/71600)] [Medline: [40835414](https://pubmed.ncbi.nlm.nih.gov/40835414/)]

66. Kilfoy A, Hsu TCC, Stockton-Powdrell C, Whelan P, Chu CH, Jibb L. An umbrella review on how digital health intervention co-design is conducted and described. *NPJ Digit Med*. Dec 23, 2024;7(1):374. [doi: [10.1038/s41746-024-01385-1](https://doi.org/10.1038/s41746-024-01385-1)] [Medline: [39715947](https://pubmed.ncbi.nlm.nih.gov/39715947/)]
67. McLeod C, Norman R, Litton E, Saville BR, Webb S, Snelling TL. Choosing primary endpoints for clinical trials of health care interventions. *Contemp Clin Trials Commun*. Dec 2019;16:100486. [doi: [10.1016/j.conctc.2019.100486](https://doi.org/10.1016/j.conctc.2019.100486)] [Medline: [31799474](https://pubmed.ncbi.nlm.nih.gov/31799474/)]
68. Yao H, Liao Z, Zhang X, et al. A comprehensive survey of the clinical trial landscape on digital therapeutics. *Heliyon*. Aug 30, 2024;10(16):e36115. [doi: [10.1016/j.heliyon.2024.e36115](https://doi.org/10.1016/j.heliyon.2024.e36115)] [Medline: [39224386](https://pubmed.ncbi.nlm.nih.gov/39224386/)]
69. Ono M, Iwasaki K. Comprehensive analysis of clinical studies and regulations of therapeutic applications in the United States and Japan. *Ther Innov Regul Sci*. Jan 2023;57(1):86-99. [doi: [10.1007/s43441-022-00442-9](https://doi.org/10.1007/s43441-022-00442-9)] [Medline: [36070067](https://pubmed.ncbi.nlm.nih.gov/36070067/)]
70. Bellucci C, Hughes K, Toomey E, Williamson PR, Matvienko-Sikar K. A survey of knowledge, perceptions and use of core outcome sets among clinical trialists. *Trials*. Dec 19, 2021;22(1):937. [doi: [10.1186/s13063-021-05891-5](https://doi.org/10.1186/s13063-021-05891-5)] [Medline: [34924001](https://pubmed.ncbi.nlm.nih.gov/34924001/)]
71. Williamson PR, Altman DG, Bagley H, et al. The COMET handbook: version 1.0. *Trials*. Jun 20, 2017;18(Suppl 3):280. [doi: [10.1186/s13063-017-1978-4](https://doi.org/10.1186/s13063-017-1978-4)] [Medline: [28681707](https://pubmed.ncbi.nlm.nih.gov/28681707/)]
72. Kirkham JJ, Gorst S, Altman DG, et al. Core Outcome Set-STAndards for Reporting: the COS-STAR statement. *PLoS Med*. Oct 2016;13(10):e1002148. [doi: [10.1371/journal.pmed.1002148](https://doi.org/10.1371/journal.pmed.1002148)] [Medline: [27755541](https://pubmed.ncbi.nlm.nih.gov/27755541/)]
73. Gargon E, Gorst SL, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 5th annual update to a systematic review of core outcome sets for research. *PLoS One*. 2019;14(12):e0225980. [doi: [10.1371/journal.pone.0225980](https://doi.org/10.1371/journal.pone.0225980)] [Medline: [31830081](https://pubmed.ncbi.nlm.nih.gov/31830081/)]
74. Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials*. Aug 6, 2012;13(1):132. [doi: [10.1186/1745-6215-13-132](https://doi.org/10.1186/1745-6215-13-132)] [Medline: [22867278](https://pubmed.ncbi.nlm.nih.gov/22867278/)]
75. Kim SY, Moon JY, Shin J, Sim JY, Kim M, Jang J. Survey for government policies regarding strategies for the commercialization and globalization of digital therapeutics. *Yonsei Med J*. Jan 2022;63(Suppl):S56-S62. [doi: [10.3349/ymj.2022.63.S56](https://doi.org/10.3349/ymj.2022.63.S56)] [Medline: [35040606](https://pubmed.ncbi.nlm.nih.gov/35040606/)]

Abbreviations

BfArM: Bundesinstitut für Arzneimittel und Medizinprodukte
COREQ: Consolidated Criteria for Reporting Qualitative Research
COS: core outcome sets
DiGA: digital health applications or Digitale Gesundheitsanwendung (German)
DTA: Digital Therapeutics Alliance
DTx: digital therapeutics
INAMI-RIZIV: National Institute for Health and Disability Insurance
MCID: minimal clinically important difference
O&M: outcome and measurement methods

Edited by Amaryllis Mavragani; peer-reviewed by Noori Kim, Sapphire Lin; submitted 24.Oct.2025; final revised version received 03.Aug.2026; accepted 07.Sep.2026; published 15.Oct.2026

Please cite as:

Mahlmann J, Mainz A, Weirauch V, Mortsiefer A, Meister S
Selection of Outcomes and Measurement Methods in Clinical Trials for Digital Therapeutics: Qualitative Analysis
JMIR Form Res 2026;10:e86437
 URL: <https://formative.jmir.org/2026/1/e86437>
 doi: [10.2196/86437](https://doi.org/10.2196/86437)

© Johannes Mahlmann, Vera Weirauch, Anne Mainz, Achim Mortsiefer, Sven Meister. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 15.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.