

Viewpoint

The Digital Therapeutics Real-World Evidence Framework: An Approach for Guiding Evidence-Based Digital Therapeutics Design, Development, Testing, and Monitoring

Meelim Kim^{1,2,3,4}, PhD; Kevin Patrick^{1,3}, MS, MD; Camille Nebeker^{1,3,4}, PhD; Job Godino^{1,3,5}, PhD; Spencer Stein⁶, MPH; Predrag Klasnja⁷, PhD; Olga Perski^{1,8}, PhD; Clare Viglione¹, MPH; Aaron Coleman⁹, BA; Eric Hekler^{1,3,4}, PhD

¹Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, United States

²Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

³The Qualcomm Institute, University of California San Diego, La Jolla, CA, United States

⁴The Design Lab, University of California San Diego, La Jolla, CA, United States

⁵Laura Rodriguez Research Institute, Family Health Centers of San Diego, San Diego, CA, United States

⁶Spiral Health Inc, San Diego, CA, United States

⁷School of Information, University of Michigan, Ann Arbor, MI, United States

⁸Faculty of Social Sciences, Tampere University, Tampere, Finland

⁹Small Steps Labs LLC dba Fitabase Inc, San Diego, CA, United States

Corresponding Author:

Eric Hekler, PhD

Herbert Wertheim School of Public Health and Human Longevity Science

University of California San Diego

9500 Gilman Dr

MC0811

La Jolla, CA, 92093

United States

Phone: 1 858 429 9370

Email: ehekler@ucsd.edu

Abstract

Digital therapeutics (DTx) are a promising way to provide safe, effective, accessible, sustainable, scalable, and equitable approaches to advance individual and population health. However, developing and deploying DTx is inherently complex in that DTx includes multiple interacting components, such as tools to support activities like medication adherence, health behavior goal-setting or self-monitoring, and algorithms that adapt the provision of these according to individual needs that may change over time. While myriad frameworks exist for different phases of DTx development, no single framework exists to guide evidence production for DTx across its full life cycle, from initial DTx development to long-term use. To fill this gap, we propose the DTx real-world evidence (RWE) framework as a pragmatic, iterative, milestone-driven approach for developing DTx. The DTx RWE framework is derived from the 4-phase development model used for behavioral interventions, but it includes key adaptations that are specific to the unique characteristics of DTx. To ensure the highest level of fidelity to the needs of users, the framework also incorporates real-world data (RWD) across the entire life cycle of DTx development and use. The DTx RWE framework is intended for any group interested in developing and deploying DTx in real-world contexts, including those in industry, health care, public health, and academia. Moreover, entities that fund research that supports the development of DTx and agencies that regulate DTx might find the DTx RWE framework useful as they endeavor to improve how DTx can advance individual and population health.

(*J Med Internet Res* 2024;26:e49208) doi: [10.2196/49208](https://doi.org/10.2196/49208)

KEYWORDS

accessible; decision making; decision; decision-based evidence-making; development; digital therapeutics; medication adherence; monitoring; pharmaceuticals; public health; real-world data; real-world evidence; safe; testing; therapeutics

Introduction

Digital therapeutics (DTx) are health software tools designed to prevent, treat, or alleviate a disease, disorder, condition, or injury by delivering interventions that have demonstrable positive therapeutic effects on individual health and produce real-world outcomes [1-3]. DTx are often complex interventions [4] as they include multiple components, such as goal-setting or problem-solving elements, and algorithms that adapt the provision of support to each person's changing needs. Common goals for DTx include improving medication adherence or regular use of medical devices (eg, glucometers), facilitating behavior change, such as improving diet, physical activity, or sleep, or improving mental health, such as care for depression, anxiety, or stress. DTx can also supplement other care, such as additional support in between clinic visits. The aspiration is that DTx can improve a patient's health outcomes, reduce the burden on health care professionals, and increase access to and usability of interventions [1,5], by providing safe, effective, and equitable support for individual and population health [6]. While myriad frameworks exist for DTx development, to date, no single unifying framework guides DTx evidence production and regulatory decision-making [7-11]. By evidence production, we mean the use of scientific methods and processes to produce meaningful data about interventions, such as DTx, both qualitative and quantitative. By regulatory decision-making, we mean the set of oversight activities governing bodies such as the US Food and Drug Administration (FDA) engage in to ensure the products or services in a targeted sector (eg, the pharmaceutical industry) are safe, effective, and aligned with individual and societal needs. We believe that a framework that streamlines linkages between evidence production and regulatory decision-making for DTx will accelerate the development, adoption, and impact of DTx.

For a comprehensive framework to be successful, it must address 2 overarching issues that distinguish DTx from other therapeutics commonly used in health care. The first issue is that DTx are large pieces of software and thus benefit from decades of experience in how software is developed, used, and improved over time. The second issue relates to the relatively new regulatory environment for DTx, which has unique demands likely to evolve further as the field of DTx advances. With respect to the first issue, the basis of DTx in software renders them as dynamic entities that benefit from, and indeed require, periodic upgrades and regular maintenance to ensure they fit with ever-evolving user needs and technological changes. DTx needs to be interoperable with the constantly changing landscape of other software solutions used within health care and requires high levels of software sophistication based on enterprise-grade code embedded within a robust system architecture that supports security, privacy, and ongoing maintenance. Research-grade code is often of good enough quality to enable a novel digital health tool to be tested in small studies and efficacy trials, such as the activities commonly done by academics when engaging in frameworks like the Obesity-Related Behavioral Intervention Trials (ORBIT) [12], but it is rarely of sufficient quality to be sustainably deployed in real-world contexts. Thus, to be successful, a new framework must be able to guide appropriate

evidence production that matches the inherent dynamic and often context-dependent nature of DTx.

The second overarching issue a new framework must address relates to regulatory issues. To date, regulation of DTx typically follows one of three approaches: (1) providing relatively limited guidance on evidence production, biasing toward the trustworthiness of DTx companies, as used in the FDA Precertification (Pre-Cert) Program [13]; (2) using emerging standards relevant to real-world data (RWD) and real-world evidence (RWE), such as reliance on data quality standards, use of RWD to efficiently run simulated clinical trials [14-17], and open science practices [18,19]; or (3) simply following variations of the 4-phase model [12] originally created for pharmaceuticals [20-22]. Payors are not providing adequate reimbursement systems, likely in part because of these issues, causing some DTx companies to declare bankruptcy [23]. We contend that a new framework that incorporates elements of these approaches may be helpful to multiple stakeholders.

Our Approach

Based on this background, we had 2 primary objectives for our proposed DTx RWE framework (hereafter called the "Framework"): (1) to create for users a decision-focused flowchart of key steps to develop DTx, with clear go/no-go milestones needed to move between phases of DTx RWE production that maps to the needs of regulatory decision-making (a point we return to in the discussion); and (2) to provide guidance on how to use RWD to develop this RWE. The 4-phase model is adapted from the National Institutes of Health (NIH)-supported ORBIT [12] model for behavioral intervention development, with phases focused on design, development, testing, and monitoring. We also considered it important that the Framework provide guidance on RWD use and evidence production in accordance with safe, timely, effective, efficient, equitable, and patient-centered (STEEEP) targets [24], as well as accessibility [25], sustainability [26], and scalability [27].

To accomplish this, we synthesized frameworks and best practice methods from relevant fields including, but not limited to, behavioral medicine, psychology, public health, medicine, human-centered design, human-computer interaction, bioinformatics, agile software development, computer engineering, health equity research, and community-based participatory research. We drew upon a range of different scoping reviews of frameworks for DTx evidence production (eg, Torous et al [28] and Lagan et al [29]). However, rather than using a scoping review process or other formalized expert consensus approach, first and foremost, we were guided by the issues summarized above because we consider them critical to the development of successful DTx yet underemphasized to date.

While we drew from many sources, the ORBIT model, an NIH-recognized approach for behavioral intervention development that uses 4 phases analogous to pharmaceutical development, was a foundational source. However, for our purposes, the ORBIT model had important limitations. It is set up to be broad and accommodate the development and testing of a wide range of different types of behavioral interventions.

This domain-specificity can make it challenging for those who may have limited familiarity with behavioral sciences to know how to use it for their specific needs. Additionally, the ORBIT model focuses on evidence production in support of the design of phase III efficacy trials. This is well-matched to studying novel interventions or the efficacy of technologies in ideal conditions, but it is very different from our goal of optimizing the development and sustainable deployment of DTx over time.

We also drew from expert consensus recommendations, including those from the World Health Organization (WHO) [30] and consensus statements from relevant workshops hosted by the NIH [31] within the United States, when creating the Framework. When necessary, the authors used first-hand knowledge based on their participation in expert consensus statements in related fields [7,32,33]; experience with both the research methods and community practices delineated in the Framework [8,32,34]; experience teaching graduate-level methods courses on topics covered in the Framework; and as innovators engaged in the development, use, and evaluation of novel methods explicitly created for digital health evidence production [35-37].

In addition, we drew on the FDA Pre-Cert Program, which was created to assess the credibility and readiness of a group to engage in DTx evidence production. The FDA Pre-Cert Program begins with an excellence appraisal, which aims to establish credibility by demonstrating the company's readiness through evaluating organizational excellence and a culture of quality [13]. Following that, the product goes through a streamlined review process to ensure a reasonable level of safety and effectiveness assurance, which leads to a decision on whether commercial distribution is approved. Once the product is on the market, they are asked to provide RWE based on RWD with a limited list of clinical trial designs in a specified period of time. The Pre-Cert Program ensures that companies have high standards of organizational excellence, that they carry out real-world monitoring of the software as it is used, and, critically, provide a mechanism that could be used to allow DTx groups to be reimbursed in some fashion while RWE production occurs.

The DTx RWE Framework

Overview

The Framework (Figure 1) is centered on a flowchart with 4 phases analogous to the ORBIT model but adapted for DTx RWE production (phase I: design; phase II: develop; phase III: test; and phase IV: monitor). Phase I activities correspond to the “double diamond” approach [38] used in human-centered design and related methods where the problem and solution specifications are delineated. Phase II activities are drawn

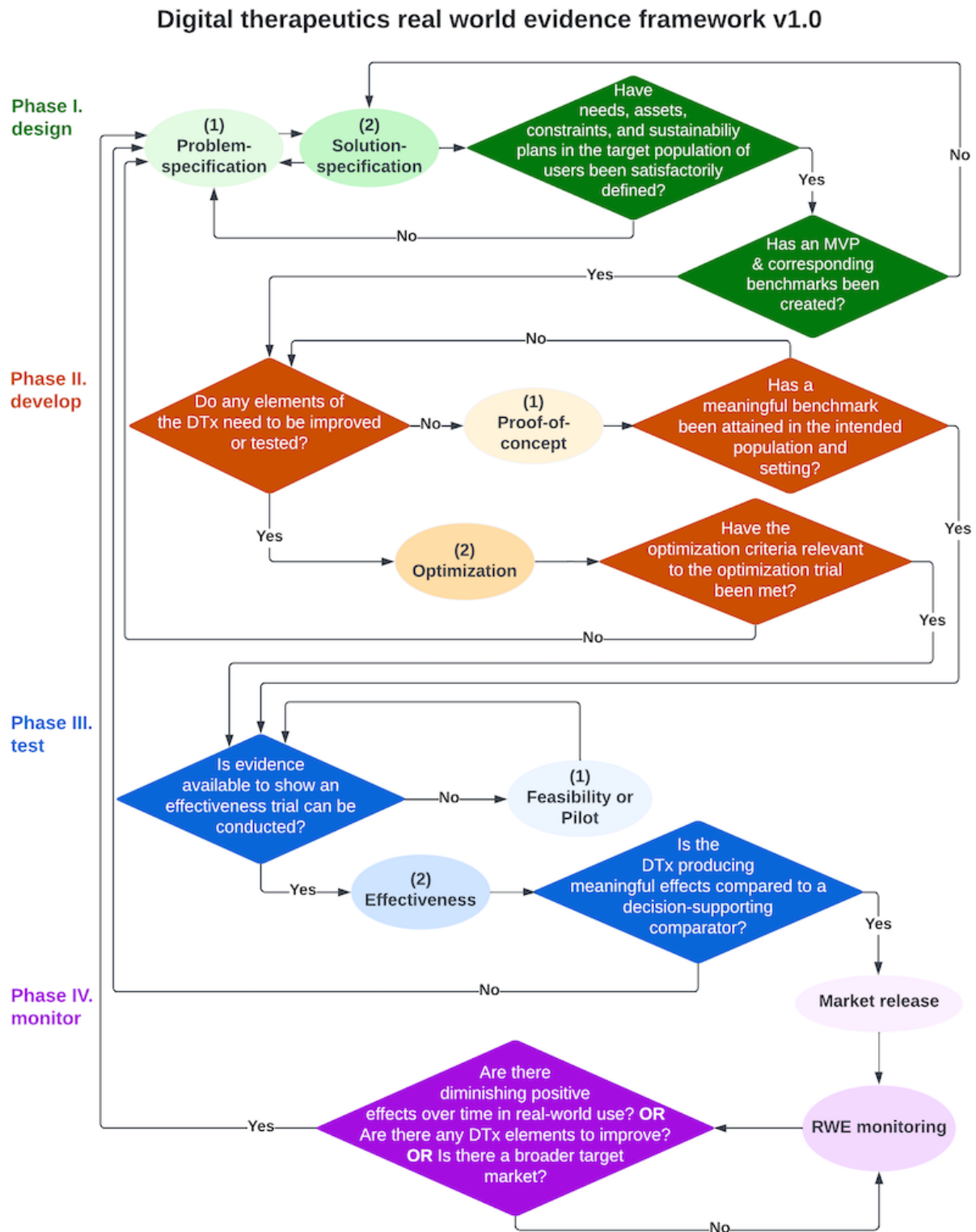
primarily from ORBIT [12], its extensions [39], and the Multiphase Optimization Strategy (MOST) [40,41]. Phase III activities are based on insights from pragmatic clinical trial best practices, including Pragmatic Explanatory Continuum Indicator Summary-2 (PRECIS-2) [42], reach effectiveness, adoption, implementation, and maintenance (RE-AIM) [43,44], and recommendations from an NIH-recognized expert panel on comparator selection for behavioral interventions [45]. Phase IV activities are drawn from implementation science [46,47], and emerging recommendations on RWE use for postmarketing surveillance from the WHO [48]. The approach to RWD in the Framework draws on recent recommendations for the use of RWD for pharmaceuticals [48-51] and on recommendations on open science best practices, which are integrated into each phase, with additional suggested recommendations [52] summarized in the Discussion section.

The Framework has 4 phases as described in detail below. Sufficient resources need to be provided at both the start of this process and as it goes through the development life cycle if it is to be successful. Moreover, analogous to what we outlined above for the FDA Pre-Cert process, we recommend that groups undertaking this process commit to the following, either on their own or through one or more partnerships, as illustrated in the Framework use case provided in [Multimedia Appendix 1](#). With this, there are 3 critical roles that must be present:

1. Designate a DTx implementor with the capability to provide ongoing, sustained deployment of the DTx. Examples could be industry, medical centers, or public health departments with proven software development and management capabilities.
2. Designate a community-serving organization that is working with and serving a target population that can provide RWD. These could be hospitals and clinics, federally qualified health centers, community clinics, or public health departments.
3. Designate a DTx evaluator with expertise in the relevant methods and approaches recommended throughout the Framework, both in terms of the flowchart of research activities and the use of RWD.

The responsibility of key stakeholders across all phases of the Framework is to intentionally consider the relevant ethical, legal, and social implications of the DTx pipeline. Including a consultant on the team who is well versed in thinking about, for example, participant characteristics and enrollment, data management (eg, collection, storage, analysis, and sharing), and related issues of bias and privacy are important throughout the process in the Framework. Engaging with an ethics review board, like a research ethics committee or institutional review board, can also be useful at various points in the process.

Figure 1. Digital therapeutics real-world evidence framework flowchart. Practical, iterative, and milestone-driven guidance to producing real-world evidence (RWE) during the design, development, testing, and monitoring phases of digital therapeutics (DTx). Oval shapes represent research or operation activities and diamond shapes represent decisions or review activities. MVP: minimal viable product.



Phase I: Design

Overview

The goal of phase I is to design the DTx product. The two activities work iteratively together: (1) problem-specification and (2) solution-specification, in accordance with the well-recognized “double diamond” used in human-centered

design [38]. Problem specification includes delineating real-world needs, constraints, assets and approaches to support future sustainability, both financial and ecological [53,54]. Solution specification focuses on iteratively creating a DTx through agile development and prototyping from low-fidelity (eg, paper concepts and storyboards) to high-fidelity (fully functional code) prototypes, with this iterative work being

situated within ethical practices and exploring the potential for reusing components or functionalities from available DTx where possible, in line with open science practices [55].

Intervention plausibility claims are the key focus of phases I and II. We are explicitly using the word “plausible” instead of “feasible” given emerging nomenclature recommendations. By “intervention plausibility,” we are referring to context-dependent probabilistic claims regarding the interaction between the DTx and targeted populations and settings, such as acceptability, demand, capacity to be adapted to a local context, etc. (a full list of plausibility targets is available at [56]; note: it was labeled feasibility in this paper given that it was published before emerging nomenclature recommendations). Progress can be determined in phase I through the specification of benchmarks that conform with the specific, measurable, actionable, realistic, and timely (SMART) goal concept [57] but are applied to establishing benchmarks. The transition from phase I to II is justifiable when benchmarks relevant to, at a minimum, safety, and effectiveness are defined and a minimal viable product (MVP) DTx has been produced that meets basic ethical, functional, and usability requirements.

Question 1a: Have Needs, Assets, Constraints, and Sustainability Plans in the Target Population of Users Been Satisfactorily Defined?

This work specifies the target population and setting and answers other key questions relevant to DTx [7]. Foundational to this is the RWD from a community-serving organization to identify unmet needs, current assets (eg, standard practices and billable activities related to the targeted DTx), and constraints (eg, likely number of billable sessions and the scope of real-world needs, if scaled). Beyond RWD, a mixed methods approach is recommended for this stage. This includes formative research such as ethnographic studies, focus groups, and interviews; community-based participatory and community-driven methods; literature reviews, such as reviewing previous intervention and epidemiological studies; and market research and analysis. Determination of whether these have been satisfactorily defined can be achieved by reaching a consensus to determine if there is overarching agreement among stakeholders, which involves a decision being made when no one objects [58].

Question 1b: Have an MVP and Corresponding Benchmarks Been Created?

The focus is to iteratively build the DTx and finalize benchmarks. In terms of DTx creation, movement to phase II is justified when a group has produced evidence showing that an MVP is functioning according to minimal usability and accessibility requirements and meets a threshold of being plausible as a tool to enable phase II activities. This could be demonstrated through DTx that are free of “bugs” and meet basic usability requirements (eg, good System Usability Score [59]). Ideally, benchmarks are established in relation to STEEEP targets [24] as well as accessibility [25], sustainability [26], and scalability [27]. To guide future implementation, data could be gathered about critical implementation issues [56]. Minimal benchmarks are needed for safety and effectiveness.

When setting benchmarks, the team should balance what is meaningful relative to current best practices with what is plausible to achieve for the target population and setting [60]. These benchmarks should, ideally, be based on RWD and establish a threshold that defines if the proposed DTx could plausibly produce desired effects safely relative to current practices. Examples of benchmarks include “a decrease of 3% in hemoglobin A_{1c} among 60% of our DTx users after 6 months of intervention” for effectiveness and “less than 10% of our DTx users experienced nonserious adverse events associated with digital treatment after a month of intervention” for safety. With a functioning MVP and benchmarks defined, the team has produced the requisite information to transition to phase II. Key approaches here include agile or lean development practices, prototype testing and development, rapid prototype testing, and qualitative methods [61]. See [Multimedia Appendix 1](#) for an example.

Phase II: Develop

Overview

The goal of phase II is to guide DTx development and optimization. There are two types of activities: (1) proof-of-concept trials and (2) optimization trials. Movement from phase II to phase III requires RWE—produced either from a proof-of-concept trial or optimization trial—that demonstrates it is plausible that the DTx can produce clinically meaningful effects in the targeted population and setting while meeting requisite implementation requirements. Given that ORBIT is our primary starting point, it is important to flag that we are rearranging where these trials take place. Specifically, in ORBIT, optimization trials occur late in phase I (ie, phase Ib). In contrast, proof-of-concept trials occur in phase IIa in ORBIT. In the authors’ opinion, the approach used in ORBIT creates 2 issues. First, it places many important activities as phase I. Second, it implicitly signals that phase I and II activities are subservient to phase III activities (a point we return to when discussing phase III of the Framework). In the Framework, we unpack activities in phase I of ORBIT to spread them across phases I and II and provide more explicit labels of the key purposes for each phase, namely design and development. With this, our intent was to provide clearer guidance on how one progresses between phases and also to connote the unique value of each phase without any need for one to be subservient to another; instead, the phase is selected based on the type of evidence production needed.

Within this phase, RWD should be used to support targeted recruitment and selection of study participants with a particular eye toward accounting for health equity in defining a target population (eg, targeting a population that explicitly underuses the current standard of care). In addition, RWD can be used to monitor for unintended consequences, both positive and negative, of the use of the DTx within the development trials. RWD could also be used, particularly with deployed DTx, for conducting data-driven algorithm development [62,63].

Question 2a: Do Any Elements of the DTx Need to Be Improved or Tested?

The proof-of-concept trial focuses primarily on intervention plausibility testing about if the overall package is producing meaningful results, as defined through the benchmarks established in phase I. Optimization trials produce evidence to improve elements of the DTx. For example, optimization trials can be used to support the evidence-based selection of DTx components (factorial trials) [64-68], refinement of components, particularly those used across time (microrandomized trials) [35,69,70], and refinement of adaptation algorithms to match the provision of support to context, individual differences, and timing (microrandomized trials [35,71], and sequential multiple assignment randomized trials [72-75]). Note also that optimization trials could also be conducted that are explicitly used to support algorithm development (eg, system identification experiments [76-78] or more data-driven algorithm development from RWD [79,80]). Like phase I, proof-of-concept and optimization trials can be used iteratively. Determining if optimization is needed is based on whether any element of a DTx needs to be improved. If no elements of the DTx package need to be refined, then a proof-of-concept trial is appropriate. If some element of the DTx package needs to be tested or improved, then an optimization trial is needed.

Question 2b: Has a Meaningful Benchmark Been Attained in the Intended Population and Setting?

This is a fundamental question for the proof-of-concept trial, an emerging approach in behavioral trials that tests a DTx in a small group (eg, 10-20) in relation to a benchmark (eg, 70% of the 10 patients shift from hypertensive to systolic blood pressure <120 mm Hg). This approach is used because of 2 known issues with small sample trials within formative work. First, small samples render the use of frequentist inferential statistics problematic [60,81,82]. Second, humans have confirmation bias, which refers to the tendency for individuals to seek out or interpret evidence in ways that align with previous beliefs, expectations, or hopes [83]. A proof-of-concept trial overcomes these challenges without the need for running larger trials, using a clearly specified a priori benchmark that can be tested using descriptive statistics. Its use of benchmarks enables resource-efficient studies with clear go/no go decision-making that reduces the risk of falling prey to confirmatory bias. Quasi-experimental, single-case, or within-participant designs in which the participants serve as their own controls are appropriate design options for a proof-of-concept trial. Further, mixed methods, where both qualitative and quantitative data are relevant to the goals of evidence production and real-world intervention implementation plausibility [56] targets, should be used. Proof-of-concept trials provide clear go/no go milestones established a priori, thus reducing the risk of continuing when not justified, which is common with more traditional piloting study approaches [39]. If benchmarks are met, the group can either shift to an optimization trial or transition to phase III. If the benchmarks are not met, then the team should consider returning to phase I or focusing on DTx optimization.

Question 2c: Have the Optimization Criteria Relevant to the Optimization Trial Been Met?

Optimization is a concept drawn from engineering that emphasizes data-driven improvement for a DTx [41,84]. Optimization supports any problem arising with the DTx, such as the DTx costing too much, not being sufficiently adhered to, being difficult to implement in real-world contexts, being inaccessible to the target population, etc. Like a benchmark, the key logic here is to specify clear optimization criteria, meaning a definition of success that can be tested using an appropriately selected optimization trial. Specification of optimization criteria is a central focus of MOST [41,84]. The goal is to create optimization criteria that are measurable and, ideally, account for real-world constraints. For example, one could establish optimization criteria that a DTx includes only intervention components with demonstrated effectiveness, that the DTx can be deployed for under US \$50 per client, or that total interactions per week with the DTx stay below 30 minutes. These criteria can be translated into clear go/no go criteria that can be assessed in an optimization trial. If the optimization criteria are met, then this can often justify movement to phase III. Plausible optimization trials for this phase could include but are not limited to: A/B testing (as used in the technology industry for improving usability) [85-87], factorial trials as used in MOST [64,65,67,68], sequential multiple assigned randomized trials [72-75], microrandomized trials [35,36,69-71], system identification experiments [76-78], studies explicitly designed to support algorithm development [79,80], and control optimization trials [37,88]. Nahum-Shani et al [40] provide guidance on when to use common optimization trial designs. See [Multimedia Appendix 1](#) for an example.

Phase III: Test**Overview**

The goal of phase III is to test if the DTx produces meaningful improvements relative to a comparator in real-world contexts. There may be two types of activities: (1) feasibility or pilot studies; or (2) an effectiveness trial. As in phase II, we have shifted phase labeling from the original ORBIT model while still honoring the types of evidence production ORBIT generally advocates for. Specifically, in the ORBIT model, feasibility or pilot studies are conducted in phase IIb. In ORBIT, phase III is reserved purely for an efficacy trial to test if the intervention impacts health outcomes. We relabeled each phase intentionally in the Framework to allow each phase of work to be conducted and produce insights that are valuable alone, with no phase treated as subservient to other phases. Thus, we sought to have development phase activities stand alone in terms of their unique value for evidence production. We recommend this shift in thinking to clearly flag the critical, independent importance of each phase of work, particularly phase II development, which could feasibly be used in perpetuity alone as a rigorous approach to continuous quality improvement. While speculative, we contend that ORBIT and related evidence production models that implicitly or explicitly treat earlier phases as subservient to phase III trials, particularly efficacy trials in ideal conditions, send a message that privileges one type of evidence at the expense of other evidence. This is problematic, as the evidence

from the other phases is particularly important for fostering real-world implementation and health equity. Further, privileging one type of evidence over others establishes the risk of a mono-method bias within scientific knowledge. This creates issues with fostering trustworthy scientific knowledge [89], reducing confidence in any consensus statements from overemphasizing one particular type of evidence, and might slow the pace of learning and progress [90]. This is particularly true when evidence production privileges tests occurring in ideal conditions, which is a valuable focus for novel interventions. When groups are developing novel interventions, they should use the ORBIT model, given its emphasis on ensuring appropriate evidence is produced to complete a high-quality efficacy trial. With phase III of the Framework, our focus is on testing if a DTx package produces meaningful results in real-world contexts to foster evidence production with high ecological validity. Thus, in the Framework we bias toward pragmatism through a focus on effectiveness trials, inclusion of benchmarks added to a modified CONSORT (Consolidated Standards of Reporting Trials) diagram guided by RE-AIM [91] to justify generalization claims, use of decision-oriented comparator selection, and emerging best practices on power calculations [39]. Moreover, the use of hybrid clinical trials, which incorporate the focus on testing both effectiveness and implementation outcomes [92,93] can be another option to be used in phase III testing of the Framework.

Even within the context of RWE, we recognize the classic tension between pragmatism and explanatory knowledge, as illustrated in PRECIS-2 [42] (eg, recruitment, eligibility, and setting). With context-invariant interventions, such as vaccines, explanatory knowledge tends to be highly valuable for producing both robust internal validity and generalizable knowledge. Further, with highly novel interventions, explanatory knowledge is valuable to determine if a signal is present in ideal conditions for the novel intervention. For evidence production to guide regulation of use in the real-world context of DTx, we suggest a bias toward pragmatism over explanatory knowledge to support ecologically valid knowledge as an approach to increase the likelihood of generalizable knowledge. Recognizing that, just like in traditional trials, this tension must be balanced in each trial conducted based on the goals of the work and what is already known. RWD should be used to support recruitment, with a particular eye toward accounting for health equity in recruiting truly representative samples. Confidence in any claims of representation and, thus, generalizability, from the trial can be supported by clearly defining meaningful while also achievable benchmarks for reach and adoption, which can be measured and reported in a modified CONSORT diagram [91]. Furthermore, once a DTx is widely deployed, RWD could be used to run simulated clinical trials to test effectiveness in real-world contexts [14-17].

Question 3a: Is Evidence Available to Show an Effectiveness Trial Can Be Conducted?

This is a key question for feasibility or pilot studies, which are used to pave the way for a future effectiveness trial. A feasibility study examines whether and how a proposed or planned effectiveness trial can be done, but without a requirement for resemblance to the future trial [39]. Note, we are explicitly using

the word “feasibility” only to refer to attributes of a targeted future fully powered trial, in alignment with 2010 CONSORT recommendations on clinical trial nomenclature [94,95]. The word “feasibility” is sometimes used in reference to issues about the intervention, such as if it would be acceptable, have sufficient demand, or could be integrated into real-world contexts [56]. To avoid confusion, we refer to these targets as intervention plausibility. Emerging recommendations from scientific groups, such as the 2010 CONSORT recommendations and others more relevant to digital health [96,97], suggest that the term “feasibility” be used to describe the probability that a particular type of study, such as a phase III clinical trial, can be conducted with sufficient rigor and fidelity by an investigative team in a given context. Based on this, we will honor this emerging consensus and only use the term “feasibility” to refer to context-dependent probabilistic claims in relation to the likelihood that a targeted study can be conducted in a particular setting by a particular group with sufficient fidelity to allow conclusions to be drawn from it. Similarly, we use emerging naming conventions and reserve the term “pilot” to refer to a specific type of feasibility trial that implements the exact eventual full protocol of a fully powered trial, but with fewer participants. The goal of a pilot study, thus, is to gather information about the likelihood that if a full trial were conducted, the data quality would be sufficient to enable trustworthy inferences [39]. If the evidence is available to show the feasibility of an effectiveness trial, the team can proceed directly to the effectiveness trial. If this evidence is not available, then the team should consider conducting a feasibility or pilot study before effectiveness trials to increase the likelihood that, if a trial is run, it will be conducted with sufficient fidelity to provide sufficient quality evidence to guide decisions.

Question 3b: Is the DTx Producing Meaningful Effects Compared to a Decision-Supporting Comparator?

The complexities of DTx establish higher evidentiary standards for generalization claims. By generalization, we specifically focus on transportability, meaning the degree to which insights gleaned from a given study sample are relevant to a stated population and setting. Recent studies of health behaviors, cognitive processes, and emotions, all factors that may influence the effectiveness of a DTx, show that behaviors, cognitions, and emotions have multiple contextual influences, can differ widely from person to person, and fluctuate over time within individuals [50,98-103]. Thus, research studies intended to produce generalizable knowledge about the effectiveness of DTx, including where, for whom, and when a given DTx is useful, need to take these factors into account. Given these complexities with regard to DTx evidence production, transportability is difficult to establish. We suggest, first, the use of a modified CONSORT diagram that integrates insights from the RE-AIM framework, and second, benchmarks be established relevant to the percentage of plausible settings and participants enrolled and completing the study. This modified CONSORT diagram provides an approach to quantifying the degree to which a sample may or may not be representative of a stated population and setting. If, for example, there is a large disparity between an eligible population and the number of patients who are enrolled, then transportability claims would

be questionable. As with other phases in this Framework, it is suggested that achievable benchmarks related to the modified CONSORT diagram be specified a priori (eg, 80% of eligible clinics will take part, 80% of eligible staff will take part, and 50% of eligible participants will be enrolled). As before, these benchmarks need to balance the need to be ambitious while also being achievable based on what is known. Only if the benchmarks are met can transportability claims be justified. These benchmarks on the modified CONSORT diagram should be derived from RWD.

As this is an effectiveness trial, comparator selection should support real-world decision-making. For example, if the clinical or community partner has a current standard of care and they are considering replacing it with the DTx, the standard of care should be the comparator. Alternatively, if the DTx would fulfill a new area of need, a stepped wedge trial [104,105], in which the DTx is released in a phased fashion across clinics, could be considered. For detailed guidance on comparator selection, see NIH expert panel recommendations [45]. For DTx testing, options include but are not limited to between-person RCTs [106,107], including remote RCTs [108,109], cluster RCTs [110,111], stepped wedge trial [104,105], and, when sufficient RWD is available, the use of simulated clinical trials [14-17] could be considered.

We also recommend the use of best-practice recommendations for power calculations that specifically do not rely upon underpowered studies to infer effect sizes [60,112,113]. Instead, what is recommended is to establish 2 effect size estimates, a threshold of clinical significance, and a plausible effect size that could be observed in the trial. The threshold of clinical significance is the smallest effect size of interest [114] that would influence clinical decision-making based on an explicit qualitative determination of a noticeable difference. This threshold of clinical significance should be informed by RWD and can be translated from the benchmarks for effectiveness defined in phase I. The plausible effect size is the most likely effect size to be observed if the trial were conducted. This plausible effect size can be informed, in part, from the results of the proof-of-concept trial, particularly if the benchmark that was met is well matched to the threshold of clinical significance. That said, given the unreliability of small sample sizes, RWD and effect sizes from previous trials most like the proposed study should be used to establish the plausible effect size. If the plausible effect size is at or above the threshold of clinical significance, then a trial is warranted. If the plausible effect size is below the threshold of clinical significance, then the trial should not be conducted, as the results gathered would not be sufficient to make a convincing argument to change clinical practice.

If benchmarks set to the modified CONSORT diagram are not met or there is no clinically significant difference observed between the DTx and comparator, then returning to phase I or phase II activities is appropriate. If benchmark and clinically meaningful differences are observed, then results can be submitted to regulators for official review. If the trial was done ethically and responsibly and the results are positive, then regulating bodies can certify the DTx and allow the DTx to market to the population and setting that was studied within the

phase III trial. See [Multimedia Appendix 1](#) for an illustrative example.

Phase IV: Monitor

Overview

The goal of phase IV is to monitor the use of the DTx within the real world, enable DTx implementers to improve the DTx with additional RWD collected from their clinical or community partners and support the expansion of the target market for the DTx through stepwise additional assessments. This is analogous to traditional non-DTx phase IV activities, including RWD use with pharmaceuticals [48].

Question 4: Are There Diminishing Positive Effects Over Time in Real-World Use? Are There Any DTx Elements to Improve? Is There a Broader Target Market?

As described earlier and elsewhere [7,32,115], continuous improvements in the DTx are not only desired but required for a DTx. For example, user interface expectations of technologies and the use of application programming interfaces drive the evolution of technology. A web application designed and tested in the 1990s [116,117], if it was not continually updated to meet changing user expectations and remain up-to-date with related application programming interfaces, would, at best, be perceived as “old” and, at worst, would not work. Thus, prespecified quality control methods for these updates would need to be used, and this could be one area where regulatory guidelines could be helpful. For example, there is active discussion about when the accumulative changes to a DTx warrant running another clinical trial [118].

Given this, any notion of a “definitive” clinical trial, a concept traditionally used, is inappropriate for DTx. Pragmatic ways of gleaning insights about when DTx is meeting expectations across the STEEEP and related criteria listed earlier are critical to monitor over time. To support this, the implementation of science practices, particularly strategies for ongoing monitoring, thoughtful adaptation, and guidance on rigorous continuous quality improvement, can be gleaned from the dynamic sustainability framework [46]. Through ongoing monitoring, issues of potential diminishing benefit (labeled voltage drop) can be observed and used to inspire a response [22]. For example, if effectiveness levels go below some predefined threshold, regulators could provide DTx with a time-limited window for continued marketing while also requiring the DTx company to reestablish a partnership with a clinical or community partner and reengage with earlier stages of the process. With this potential risk looming, it could establish an incentive for the DTx company to engage in continuous improvement, guided by the other 2 questions, and to maintain mutually beneficial partnerships.

These recommendations conform with recommendations from the WHO for monitoring and evaluating digital health interventions [119]. According to the WHO, the 4 major components of digital health monitoring (ie, functionality, stability, fidelity, and quality) should guide ongoing monitoring, with these questions mapping onto our proposed questions:

- Are there diminishing positive effects over time in real-world use?
 - (Quality) Is the content and the delivery of the intervention of high enough quality to yield intended outcomes?
 - (Quality) How well and consistently is the intervention delivered?
- Are there any DTx elements to improve?
 - (Functionality) Does the system operate as intended?
 - (Stability) Does the system consistently operate as intended?
- Is there a broader target market?
 - (Fidelity) Do the realities of field implementation alter the functionality and stability of the system, changing the intervention from that which was intended?

RWE for postmarket surveillance is being explored, and opportunities and pitfalls that are also relevant to DTx are being articulated and should be considered regarding DTx regulation [48]. Monitoring could require benchmarks to be set for all key targets of evidence production (eg, effectiveness, safety, and equity) as one pathway for cultivating more rigor in monitoring efforts and reducing the risk of confirmation bias during phase IV.

If the DTx implementer believes their DTx can support a more diverse market share than what was approved in phase III, RWD collected in phase IV may help the DTx implementer accelerate this expansion. For example, monitoring could be used to identify plausible new populations, settings, or areas for improvement of the DTx, particularly if done with other community-serving organization partners who may have providers prescribing the DTx for “off-label” uses. One plausible way to improve evidence production during phase IV monitoring would be to focus evidence production more on testing and improving the elements of a DTx (eg, intervention components and adaptation algorithms) instead of the DTx package. A more detailed rationale for this is described elsewhere [7,32,47,82,120]. A second opportunity would be to link activities and efforts with ongoing behavioral ontology efforts to foster better knowledge comparison across various DTx [121,122]. With that said, standards for ongoing monitoring of RWD and RWE are rapidly evolving; thus, this is a critical area for continued work. See [Multimedia Appendix 1](#) for an illustrative example.

Discussion

Overview

The Framework provides guidance to groups seeking to sustainably deploy DTx for use in real-world contexts and may be helpful to regulatory and funding entities as they provide support and oversight of DTx. We acknowledge that the Framework has not been rigorously vetted and that additional work is needed to establish its value. This includes determining whether the use of the Framework has greater or lesser utility in specific domains of DTx applications, such as those used in mental health, behavioral health, or as an adjunct to

pharmacological and other interventions for chronic diseases like cancer, musculoskeletal disorders, and cardiovascular disease. We know of no specific reasons why such differences should exist, but as published reports of DTx research emerge in the future, these distinctions might become evident.

Another issue pertains to how the Framework can assist with evaluating DTx that are already in the field, including digital wellness tools that do not meet the definition of a DTx. The longer a DTx has been sustainably deployed at scale, the more likely it is that simulated clinical trial methods could be used to study DTx along STEEEP criteria. With this, efficiencies could be further advanced for evidence production through simulated clinical trials through RWD. Future work would benefit from continued focus on the refined development of simulated clinical trial best practices to improve the pace and resource efficiency of learning.

Regarding digital health wellness tools, these tools often build on foundational behavior change techniques, such as self-monitoring and goal setting that have decades of evidence to guide their design and implementation. When situations like this exist, the burden of proof should be to justify why current evidence is not sufficient already. The most likely gaps in research for these may relate to insufficient evidence for their effectiveness in a broad range of users or settings, so a targeted adaptation of the Framework to fill in these gaps might be the best approach. For example, future work could explore ways to adapt the Framework for use with community-based organizations and community-serving well-being institutions such as the YMCA or Jewish Family Services, along with corporate wellness and related wellness programs that are not implemented by or in partnership with the health care system.

With these limitations recognized, we expand below on the need for three areas of future work related to the Framework: (1) how using RWD advances health equity; (2) cultivating trustworthy partnerships that foster the use of Framework, as a secondary pathway to advance health equity; and (3) suggesting next steps with regard to regulation and funding.

Advancing Health Equity Through RWD

Overview

In our view, increased sophistication on the effective use of RWD can become a critical tool to overcome some of the major challenges currently faced in health care, including identifying and addressing health disparities to advance health equity for all, and to foster more targeted and resource-efficient evidence production. RWD provides the information needed to specify unmet needs in general as well as those for individuals, communities, and populations where current practices are not producing desired results. This can help focus resource expenditures and efforts to reduce health disparities.

Future work could advance the use of RWD to drive the development of evidence-based solutions that serve communities most in need. Clinically meaningful benchmarks based on RWD provide an approach for guiding DTx development, both for individual DTxs and for DTxs at large. These would create pressure for DTx not simply to replicate existing standards of care but to improve upon them. Indeed, RWD can be used to

establish benchmarks across the various evidence production targets, such as effectiveness, safety, and equity, to provide the foundational data needed to measure, monitor, and, thus, drive equitable progress in individual and population health.

Cultivating Trustworthy Partnerships

As presented in the section above introducing the Framework, we recommend a tripartite approach to its use, comprising an entity committed to sustaining the DTx, a community-serving organization from which the RWD comes, and an entity with appropriate expertise in RWE evaluation efforts. While these conditions can be met within well-resourced settings such as academic medical centers, we suggest that there are opportunities for implementing the Framework through partnerships among groups that may historically not have worked as closely together, such as industry partners working with federally qualified health centers and supported with an academic partner, as illustrated in [Multimedia Appendix 1](#).

The trustworthiness of all actors involved must be acknowledged as a foundational starting point for any approach to evidence production [123,124]. This includes not merely thinking that trust can be achieved with effective communication but that, at its core, trust involves acknowledging and centering ethics, inclusion, and equity as central guiding principles in the work [89,123]. To do this, we propose the use of best practices in cultivating and maintaining partnerships that have already been delineated like community-based participatory research [125,126], patient-led innovation [127,128], community-driven design [129], community psychology practices [130,131], and ethical digital health research practices [33,34,132,133]. Incorporation of approaches to determining corporate trustworthiness that was formatively tested in the FDA Pre-Cert program can be used, including excellence appraisal, and streamlined review elements (eg, real-world performance plan and review determination information) [13]. The Digital Health

Checklist [133] might also be helpful to guide ethical practices for evidence production relevant to DTx pertaining to issues such as accessibility, privacy, data management, balancing risks, and benefits, all grounded in fundamental ethical principles including respect for persons, beneficence, justice, and respect for law and the public interest.

Regulatory and Funding Issues

We encourage regulators and funders of DTx to explore whether the Framework can help guide their efforts. The principles embodied in the Framework could be used to establish generalized regulatory expectations for DTx. Clarifying these expectations could help get multiple DTx developers and purchases “on the same page” with respect to achieving and maintaining appropriate standards of quality throughout the life cycle of DTx use. Similarly, funders of DTx research and development, such as the NIH, the Agency for Healthcare Research and Quality, the Patient-Centered Outcomes Research Institute, and the Health Resources and Services Administration, could encourage applicants to use the Framework, and if they do, then demonstrate how they propose to achieve the benchmarks that it includes.

Conclusion

The Framework is intended to improve evidence production and sustainable deployment of DTx in real-world contexts. The Framework provides guidance on how to design, develop, test, and monitor DTx, both in the early stages of their development and over time as they are used in real-world contexts. Our hope is that the Framework can help address issues commonly seen with DTx, including low DTx uptake, long-term sustainability, and insufficient attention to health disparities. Overall, there is considerable opportunity to improve individual and population health equitably through DTx, and we hope the Framework can contribute to this end.

Acknowledgments

We want to thank Dr Kenneth Freedland for providing expert consultation and review of earlier versions of this manuscript. This study is supported by the National Library of Medicine (R01LM013107) and the National Cancer Institute (R01CA244777) of the National Institute of Health. CN was supported by the Patient-Centered Outcomes Research Institute (PCORI; award ME-2020C3-21310) and the Altman Clinical and Translational Research Institute (ACTRI) at the University of California, San Diego. The ACTRI is funded from awards issued by the National Center for Advancing Translational Sciences (NIH UL1TR001442). PK is supported by the National Cancer Institute (1U01CA229445) of the National Institute of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or PCORI.

Authors' Contributions

MK, KP, CN, and EH contributed to conceptualizing, drafting, reviewing, and finalizing the paper. All authors contributed to critical reviews and revisions of the paper and have given approval for the paper to be submitted for publication. The authors read and approved the final manuscript.

Conflicts of Interest

SS has an ownership interest in Spiral Health, Inc. AC has an ownership interest in Fitabase, Inc. OP acts as an unpaid scientific advisor for the Smoke Free app. All other authors declare no competing interests.

Multimedia Appendix 1

The hypothetical example of DTx RWE Framework.

[\[DOCX File, 34 KB-Multimedia Appendix 1\]](#)

References

1. Understanding DTx: a new category of medicine. Digital Therapeutics Alliance. 2022. URL: <https://dtxalliance.org/understanding-dtx/> [accessed 2024-02-02]
2. Digital therapeutics get a brand new definition. Healthcare Brew. 2023. URL: <https://tinyurl.com/y4uyr9ww> [accessed 2024-02-02]
3. ISO/TR 11147:2023(en) health informatics—personalized digital health—digital therapeutics health software systems. International Organization for Standardization (ISO). 2023. URL: <https://tinyurl.com/bddx8vtx> [accessed 2024-02-02]
4. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337:a1655. [FREE Full text] [doi: [10.1136/bmj.a1655](https://doi.org/10.1136/bmj.a1655)] [Medline: [18824488](https://pubmed.ncbi.nlm.nih.gov/18824488/)]
5. Patel NA, Butte AJ. Characteristics and challenges of the clinical pipeline of digital therapeutics. *NPJ Digit Med*. 2020;3(1):159. [FREE Full text] [doi: [10.1038/s41746-020-00370-8](https://doi.org/10.1038/s41746-020-00370-8)] [Medline: [33311567](https://pubmed.ncbi.nlm.nih.gov/33311567/)]
6. Dang A, Arora D, Rane P. Role of digital therapeutics and the changing future of healthcare. *J Family Med Prim Care*. 2020;9(5):2207-2213. [FREE Full text] [doi: [10.4103/jfmpc.jfmpc_105_20](https://doi.org/10.4103/jfmpc.jfmpc_105_20)] [Medline: [32754475](https://pubmed.ncbi.nlm.nih.gov/32754475/)]
7. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. *Am J Prev Med*. 2016;51(5):843-851. [FREE Full text] [doi: [10.1016/j.amepre.2016.06.008](https://doi.org/10.1016/j.amepre.2016.06.008)] [Medline: [27745684](https://pubmed.ncbi.nlm.nih.gov/27745684/)]
8. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *J Med Internet Res*. 2017;19(6):e232. [FREE Full text] [doi: [10.2196/jmir.7126](https://doi.org/10.2196/jmir.7126)] [Medline: [28663162](https://pubmed.ncbi.nlm.nih.gov/28663162/)]
9. Voorheis P, Zhao A, Kuluski K, Pham Q, Scott T, Sztur P, et al. Integrating behavioral science and design thinking to develop mobile health interventions: systematic scoping review. *JMIR Mhealth Uhealth*. 2022;10(3):e35799. [FREE Full text] [doi: [10.2196/35799](https://doi.org/10.2196/35799)] [Medline: [35293871](https://pubmed.ncbi.nlm.nih.gov/35293871/)]
10. Torous J. Evaluating apps: making informed decisions with available real-world data. Digital health interventions from wellness to therapeutics: development and dissemination. 2022. Presented at: National Institutes of Health, Digital Health Intervention Workshop; July 13, 2022, 2022; NA.
11. Miao BY, Arneson D, Wang M, Butte AJ. Open challenges in developing digital therapeutics in the United States. *PLOS Digit Health*. 2022;1(1):e0000008. [FREE Full text] [doi: [10.1371/journal.pdig.0000008](https://doi.org/10.1371/journal.pdig.0000008)] [Medline: [36812515](https://pubmed.ncbi.nlm.nih.gov/36812515/)]
12. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. *Health Psychol*. 2015;34(10):971-982. [doi: [10.1037/hea0000161](https://doi.org/10.1037/hea0000161)] [Medline: [25642841](https://pubmed.ncbi.nlm.nih.gov/25642841/)]
13. The Software Precertification (Pre-Cert) pilot program: tailored total product lifecycle approaches and key findings. U.S. Food & Drug Administration. 2022. URL: <https://tinyurl.com/ycwsjh8u> [accessed 2024-02-02]
14. Chen Z, Zhang H, Guo Y, George TJ, Prosperi M, Hogan WR, et al. Exploring the feasibility of using real-world data from a large clinical data research network to simulate clinical trials of Alzheimer's disease. *NPJ Digit Med*. 2021;4(1):84. [FREE Full text] [doi: [10.1038/s41746-021-00452-1](https://doi.org/10.1038/s41746-021-00452-1)] [Medline: [33990663](https://pubmed.ncbi.nlm.nih.gov/33990663/)]
15. Swift B, Jain L, White C, Chandrasekaran V, Bhandari A, Hughes DA, et al. Innovation at the intersection of clinical trials and real-world data science to advance patient care. *Clin Transl Sci*. 2018;11(5):450-460. [FREE Full text] [doi: [10.1111/cts.12559](https://doi.org/10.1111/cts.12559)] [Medline: [29768712](https://pubmed.ncbi.nlm.nih.gov/29768712/)]
16. Müller P, Chandra NK, Sarkar A. Bayesian approaches to include real-world data in clinical studies. *Philos Trans A Math Phys Eng Sci*. 2023;381(2247):20220158. [FREE Full text] [doi: [10.1098/rsta.2022.0158](https://doi.org/10.1098/rsta.2022.0158)] [Medline: [36970825](https://pubmed.ncbi.nlm.nih.gov/36970825/)]
17. Wang C, Rosner GL. A bayesian nonparametric causal inference model for synthesizing randomized clinical trial and real-world evidence. *Stat Med*. 2019;38(14):2573-2588. [doi: [10.1002/sim.8134](https://doi.org/10.1002/sim.8134)] [Medline: [30883861](https://pubmed.ncbi.nlm.nih.gov/30883861/)]
18. Kokol P. Agile software development in healthcare: a synthetic scoping review. *Appl Sci*. 2022;12(19):9462. [FREE Full text] [doi: [10.3390/app12199462](https://doi.org/10.3390/app12199462)]
19. Van Velthoven MH, Smith J, Wells G, Brindley D. Digital health app development standards: a systematic review protocol. *BMJ Open*. 2018;8(8):e022969. [FREE Full text] [doi: [10.1136/bmjopen-2018-022969](https://doi.org/10.1136/bmjopen-2018-022969)]
20. Basic principles on use of registries in approval applications. Pharmaceuticals and Medical Devices Agency. URL: <https://www.pmda.go.jp/files/000240810.pdf> [accessed 2024-02-02]
21. Guideline on the application of real world evidence of medical device. Ministry of Food and Drug Safety (MFDS). 2019. URL: <https://tinyurl.com/mr3zpj85> [accessed 2024-02-02]
22. Framework for FDA's real-world evidence program. U.S. Food & Drug Administration. 2018. URL: <https://www.fda.gov/media/120060/download> [accessed 2024-02-02]

23. Pear therapeutics assets sold for \$6M at auction after bankruptcy. MobiHealthNews. 2023. URL: <https://tinyurl.com/4v7b87kk> [accessed 2024-02-02]
24. Six domains of healthcare quality. Agency for Healthcare Research and Quality. 2022. URL: <https://www.ahrq.gov/talkingquality/measures/six-domains.html> [accessed 2024-02-02]
25. Gulliford M, Figueroa-Munoz J, Morgan M, Hughes D, Gibson B, Beech R, et al. What does 'access to health care' mean? *J Health Serv Res Policy*. 2002;7(3):186-188. [doi: [10.1258/135581902760082517](https://doi.org/10.1258/135581902760082517)] [Medline: [12171751](https://pubmed.ncbi.nlm.nih.gov/12171751/)]
26. McCool J, Dobson R, Muinga N, Paton C, Pagliari C, Agawal S, et al. Factors influencing the sustainability of digital health interventions in low-resource settings: lessons from five countries. *J Glob Health*. 2020;10(2):020396. [FREE Full text] [doi: [10.7189/jogh.10.020396](https://doi.org/10.7189/jogh.10.020396)] [Medline: [33274059](https://pubmed.ncbi.nlm.nih.gov/33274059/)]
27. Milat A, Lee K, Conte K, Grunseit A, Wolfenden L, van Nassau F, et al. Intervention scalability assessment tool: a decision support tool for health policy makers and implementers. *Health Res Policy Syst*. 2020;18(1):1. [FREE Full text] [doi: [10.1186/s12961-019-0494-2](https://doi.org/10.1186/s12961-019-0494-2)] [Medline: [31900230](https://pubmed.ncbi.nlm.nih.gov/31900230/)]
28. Torous JB, Chan SR, Gipson SYMT, Kim JW, Nguyen TQ, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. *Psychiatr Serv*. 2018;69(5):498-500. [FREE Full text] [doi: [10.1176/appi.ps.201700423](https://doi.org/10.1176/appi.ps.201700423)] [Medline: [29446337](https://pubmed.ncbi.nlm.nih.gov/29446337/)]
29. Lagan S, Sandler L, Torous J. Evaluating evaluation frameworks: a scoping review of frameworks for assessing health apps. *BMJ Open*. 2021;11(3):e047001. [FREE Full text] [doi: [10.1136/bmjopen-2020-047001](https://doi.org/10.1136/bmjopen-2020-047001)] [Medline: [33741674](https://pubmed.ncbi.nlm.nih.gov/33741674/)]
30. Consensus statement: role of policy-makers and health care leaders in implementation of the Global Patient Safety Action Plan 2021–2030. World Health Organization. 2022. URL: <https://www.who.int/publications/i/item/WHO-UHL-IHS-PSF-2022.3> [accessed 2024-02-02]
31. NIH consensus statement. National Institutes of Health (NIH). 2000. URL: <https://tinyurl.com/4txypc3d> [accessed 2024-02-02]
32. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The pace of technologic change: implications for digital health behavior intervention research. *Am J Prev Med*. 2016;51(5):816-824. [doi: [10.1016/j.amepre.2016.05.001](https://doi.org/10.1016/j.amepre.2016.05.001)] [Medline: [27745681](https://pubmed.ncbi.nlm.nih.gov/27745681/)]
33. Nebeker C, Torous J, Ellis RJB. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Med*. 2019;17(1):137. [FREE Full text] [doi: [10.1186/s12916-019-1377-7](https://doi.org/10.1186/s12916-019-1377-7)] [Medline: [31311535](https://pubmed.ncbi.nlm.nih.gov/31311535/)]
34. Nebeker C, Ellis RJB, Torous J. Development of a decision-making checklist tool to support technology selection in digital health research. *Transl Behav Med*. 2020;10(4):1004-1015. [FREE Full text] [doi: [10.1093/tbm/ibz074](https://doi.org/10.1093/tbm/ibz074)] [Medline: [31120511](https://pubmed.ncbi.nlm.nih.gov/31120511/)]
35. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. *Health Psychol*. 2015;34S:1220-1228. [FREE Full text] [doi: [10.1037/hea0000305](https://doi.org/10.1037/hea0000305)] [Medline: [26651463](https://pubmed.ncbi.nlm.nih.gov/26651463/)]
36. Klasnja P, Smith S, Seewald NJ, Lee A, Hall K, Luers B, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. *Ann Behav Med*. 2019;53(6):573-582. [FREE Full text] [doi: [10.1093/abm/kay067](https://doi.org/10.1093/abm/kay067)] [Medline: [30192907](https://pubmed.ncbi.nlm.nih.gov/30192907/)]
37. Hekler EB, Rivera DE, Martin CA, Phatak SS, Freigoun MT, Korinek E, et al. Tutorial for using control systems engineering to optimize adaptive mobile health interventions. *J Med Internet Res*. 2018;20(6):e214. [FREE Full text] [doi: [10.2196/jmir.8622](https://doi.org/10.2196/jmir.8622)] [Medline: [29954725](https://pubmed.ncbi.nlm.nih.gov/29954725/)]
38. Framework for innovation: Design Council's evolved double diamond. Design Council. 2019. URL: <https://tinyurl.com/bdhmmyuw> [accessed 2024-02-02]
39. Powell LH, Freedland KE, Kaufmann PG. Behavioral Clinical Trials for Chronic Diseases: Scientific Foundations. Cham, Switzerland. Springer International Publishing; 2021.
40. Nahum-Shani I, Dziak JJ, Wetter DW. MCMTTC: a pragmatic framework for selecting an experimental design to inform the development of digital interventions. *Front Digit Health*. 2022;4:798025. [FREE Full text] [doi: [10.3389/fdgth.2022.798025](https://doi.org/10.3389/fdgth.2022.798025)] [Medline: [35355685](https://pubmed.ncbi.nlm.nih.gov/35355685/)]
41. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). Cham, Switzerland. Springer International Publishing; 2018.
42. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*. 2015;350:h2147. [FREE Full text] [doi: [10.1136/bmj.h2147](https://doi.org/10.1136/bmj.h2147)] [Medline: [25956159](https://pubmed.ncbi.nlm.nih.gov/25956159/)]
43. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. *Front Public Health*. 2019;7:64. [FREE Full text] [doi: [10.3389/fpubh.2019.00064](https://doi.org/10.3389/fpubh.2019.00064)] [Medline: [30984733](https://pubmed.ncbi.nlm.nih.gov/30984733/)]
44. Holtrop JS, Estabrooks PA, Gaglio B, Harden SM, Kessler RS, King DK, et al. Understanding and applying the RE-AIM framework: clarifications and resources. *J Clin Transl Sci*. 2021;5(1):e126. [FREE Full text] [doi: [10.1017/cts.2021.789](https://doi.org/10.1017/cts.2021.789)] [Medline: [34367671](https://pubmed.ncbi.nlm.nih.gov/34367671/)]
45. Freedland KE, King AC, Ambrosius WT, Mayo-Wilson E, Mohr DC, Czajkowski SM, et al. The selection of comparators for randomized controlled trials of health-related behavioral interventions: recommendations of an NIH expert panel. *J Clin Epidemiol*. 2019;110:74-81. [FREE Full text] [doi: [10.1016/j.jclinepi.2019.02.011](https://doi.org/10.1016/j.jclinepi.2019.02.011)] [Medline: [30826377](https://pubmed.ncbi.nlm.nih.gov/30826377/)]
46. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. *Implement Sci*. 2013;8:117. [FREE Full text] [doi: [10.1186/1748-5908-8-117](https://doi.org/10.1186/1748-5908-8-117)] [Medline: [24088228](https://pubmed.ncbi.nlm.nih.gov/24088228/)]

47. Hekler EB, Klasnja P, Riley WT, Buman MP, Huberty J, Rivera DE, et al. Agile science: creating useful products for behavior change in the real world. *Transl Behav Med*. 2016;6(2):317-328. [FREE Full text] [doi: [10.1007/s13142-016-0395-7](https://doi.org/10.1007/s13142-016-0395-7)] [Medline: [27357001](https://pubmed.ncbi.nlm.nih.gov/27357001/)]
48. Beaulieu-Jones BK, Finlayson SG, Yuan W, Altman RB, Kohane IS, Prasad V, et al. Examining the use of real-world evidence in the regulatory process. *Clin Pharmacol Ther*. 2020;107(4):843-852. [FREE Full text] [doi: [10.1002/cpt.1658](https://doi.org/10.1002/cpt.1658)] [Medline: [31562770](https://pubmed.ncbi.nlm.nih.gov/31562770/)]
49. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health*. 2007;10(5):326-335. [FREE Full text] [doi: [10.1111/j.1524-4733.2007.00186.x](https://doi.org/10.1111/j.1524-4733.2007.00186.x)] [Medline: [17888097](https://pubmed.ncbi.nlm.nih.gov/17888097/)]
50. Hekler E, Tiro JA, Hunter CM, Nebeker C. Precision health: the role of the social and behavioral sciences in advancing the vision. *Ann Behav Med*. 2020;54(11):805-826. [FREE Full text] [doi: [10.1093/abm/kaaa018](https://doi.org/10.1093/abm/kaaa018)] [Medline: [32338719](https://pubmed.ncbi.nlm.nih.gov/32338719/)]
51. Vaithinathan AG, Asokan V. Public health and precision medicine share a goal. *J Evid Based Med*. 2017;10(2):76-80. [doi: [10.1111/jebm.12239](https://doi.org/10.1111/jebm.12239)] [Medline: [28276633](https://pubmed.ncbi.nlm.nih.gov/28276633/)]
52. Burns L, Roux NL, Kalesnik-Orszulak R, Christian J, Hukkelhoven M, Rockhold F, et al. Real-world evidence for regulatory decision-making: guidance from around the world. *Clin Ther*. 2022;44(3):420-437. [FREE Full text] [doi: [10.1016/j.clinthera.2022.01.012](https://doi.org/10.1016/j.clinthera.2022.01.012)] [Medline: [35181179](https://pubmed.ncbi.nlm.nih.gov/35181179/)]
53. Chevance G, Hekler EB, Efoui-Hess M, Godino J, Golaszewski N, Gualtieri L, et al. Digital health at the age of the anthropocene. *Lancet Digit Health*. 2020;2(6):e290-e291. [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30130-8](https://doi.org/10.1016/S2589-7500(20)30130-8)] [Medline: [33328121](https://pubmed.ncbi.nlm.nih.gov/33328121/)]
54. Chevance G, Fresán U, Hekler E, Edmondson D, Lloyd SJ, Ballester J, et al. Thinking health-related behaviors in a climate change context: a narrative review. *Ann Behav Med*. 2023;57(3):193-204. [FREE Full text] [doi: [10.1093/abm/kaac039](https://doi.org/10.1093/abm/kaac039)] [Medline: [35861123](https://pubmed.ncbi.nlm.nih.gov/35861123/)]
55. Kwasnicka D, Keller J, Perski O, Potthoff S, Hoor GAT, Ainsworth B, et al. White paper: open digital health—accelerating transparent and scalable health promotion and treatment. *Health Psychol Rev*. 2022;16(4):475-491. [FREE Full text] [doi: [10.1080/17437199.2022.2046482](https://doi.org/10.1080/17437199.2022.2046482)] [Medline: [35240931](https://pubmed.ncbi.nlm.nih.gov/35240931/)]
56. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. *Am J Prev Med*. 2009;36(5):452-457. [FREE Full text] [doi: [10.1016/j.amepre.2009.02.002](https://doi.org/10.1016/j.amepre.2009.02.002)] [Medline: [19362699](https://pubmed.ncbi.nlm.nih.gov/19362699/)]
57. SMART goals: a how to guide. University of California. 2016. URL: <https://tinyurl.com/mvy8cxzt> [accessed 2024-02-02]
58. Rau T. The difference between whole-group consensus and dynamic governance/sociocracy. *Sociocracy For All*. 2023. URL: <https://tinyurl.com/ysy47skf> [accessed 2024-02-02]
59. System Usability Scale (SUS). usability.gov. URL: <https://tinyurl.com/22pkunfy> [accessed 2024-02-02]
60. Freedland KE. Pilot trials in health-related behavioral intervention research: problems, solutions, and recommendations. *Health Psychol*. 2020;39(10):851-862. [doi: [10.1037/hea0000946](https://doi.org/10.1037/hea0000946)] [Medline: [32614197](https://pubmed.ncbi.nlm.nih.gov/32614197/)]
61. Design thinking 101. Nielsen Norman Group. 2016. URL: <https://www.nngroup.com/articles/design-thinking/> [accessed 2024-02-02]
62. Martín CA, Rivera DE, Hekler EB, Riley WT, Buman MP, Adams MA, et al. Development of a control-oriented model of social cognitive theory for optimized mHealth behavioral interventions. *IEEE Trans Control Syst Technol*. 2020;28(2):331-346. [FREE Full text] [doi: [10.1109/tcst.2018.2873538](https://doi.org/10.1109/tcst.2018.2873538)] [Medline: [33746479](https://pubmed.ncbi.nlm.nih.gov/33746479/)]
63. Chevance G, Golaszewski NM, Baretta D, Hekler EB, Larsen BA, Patrick K, et al. Modelling multiple health behavior change with network analyses: results from a one-year study conducted among overweight and obese adults. *J Behav Med*. 2020;43(2):254-261. [FREE Full text] [doi: [10.1007/s10865-020-00137-2](https://doi.org/10.1007/s10865-020-00137-2)] [Medline: [31997127](https://pubmed.ncbi.nlm.nih.gov/31997127/)]
64. Spring B, Pfammatter AF, Marchese SH, Stump T, Pellegrini C, McFadden HG, et al. A factorial experiment to optimize remotely delivered behavioral treatment for obesity: results of the Opt-IN study. *Obesity (Silver Spring)*. 2020;28(9):1652-1662. [FREE Full text] [doi: [10.1002/oby.22915](https://doi.org/10.1002/oby.22915)] [Medline: [32656994](https://pubmed.ncbi.nlm.nih.gov/32656994/)]
65. Thomas JG, Goldstein CM, Bond DS, Lillis J, Hekler EB, Emerson JA, et al. Evaluation of intervention components to maximize outcomes of behavioral obesity treatment delivered online: a factorial experiment following the multiphase optimization strategy framework. *Contemp Clin Trials*. 2021;100:106217. [FREE Full text] [doi: [10.1016/j.cct.2020.106217](https://doi.org/10.1016/j.cct.2020.106217)] [Medline: [33197609](https://pubmed.ncbi.nlm.nih.gov/33197609/)]
66. Lee H, Choi EH, Shin JU, Kim TG, Oh J, Shin B, et al. The impact of intervention design on user engagement in digital therapeutics research: factorial experiment With a mixed methods study. *JMIR Form Res*. 2024;8:e51225. [doi: [10.2196/51225](https://doi.org/10.2196/51225)] [Medline: [38335015](https://pubmed.ncbi.nlm.nih.gov/38335015/)]
67. Cipriani A, Barbui C. What is a factorial trial? *Epidemiol Psychiatr Sci*. 2013;22(3):213-215. [FREE Full text] [doi: [10.1017/s2045796013000231](https://doi.org/10.1017/s2045796013000231)]
68. Kip H, Da Silva MC, Bouman YHA, van Gemert-Pijnen LJEWC, Kelders SM. A self-control training app to increase self-control and reduce aggression—a full factorial design. *Internet Interv*. 2021;25:100392. [FREE Full text] [doi: [10.1016/j.invent.2021.100392](https://doi.org/10.1016/j.invent.2021.100392)] [Medline: [33996509](https://pubmed.ncbi.nlm.nih.gov/33996509/)]
69. Seewald NJ, Smith SN, Lee AJ, Klasnja P, Murphy SA. Practical considerations for data collection and management in mobile health micro-randomized trials. *Stat Biosci*. 2019;11(2):355-370. [FREE Full text] [doi: [10.1007/s12561-018-09228-w](https://doi.org/10.1007/s12561-018-09228-w)] [Medline: [31462937](https://pubmed.ncbi.nlm.nih.gov/31462937/)]

70. Bell L, Garnett C, Qian T, Perski O, Potts HWW, Williamson E. Notifications to improve engagement with an alcohol reduction app: protocol for a micro-randomized trial. *JMIR Res Protoc*. 2020;9(8):e18690. [[FREE Full text](#)] [doi: [10.2196/18690](https://doi.org/10.2196/18690)] [Medline: [32763878](https://pubmed.ncbi.nlm.nih.gov/32763878/)]
71. Battalio SL, Conroy DE, Dempsey W, Liao P, Menictas M, Murphy S, et al. Sense2Stop: a micro-randomized trial using wearable sensors to optimize a just-in-time-adaptive stress management intervention for smoking relapse prevention. *Contemp Clin Trials*. 2021;109:106534. [[FREE Full text](#)] [doi: [10.1016/j.cct.2021.106534](https://doi.org/10.1016/j.cct.2021.106534)] [Medline: [34375749](https://pubmed.ncbi.nlm.nih.gov/34375749/)]
72. Czyz EK, King CA, Prouty D, Micol VJ, Walton M, Nahum-Shani I. Adaptive intervention for prevention of adolescent suicidal behavior after hospitalization: a pilot sequential multiple assignment randomized trial. *J Child Psychol Psychiatry*. 2021;62(8):1019-1031. [[FREE Full text](#)] [doi: [10.1111/jcpp.13383](https://doi.org/10.1111/jcpp.13383)] [Medline: [33590475](https://pubmed.ncbi.nlm.nih.gov/33590475/)]
73. De Barros Gonze B, Da Costa Padovani R, Do Socorro Simoes M, Lauria V, Proença NL, Sperandio EF, et al. Use of a smartphone app to increase physical activity levels in insufficiently active adults: feasibility Sequential Multiple Assignment Randomized Trial (SMART). *JMIR Res Protoc*. 2020;9(10):e14322. [[FREE Full text](#)] [doi: [10.2196/14322](https://doi.org/10.2196/14322)] [Medline: [33094733](https://pubmed.ncbi.nlm.nih.gov/33094733/)]
74. Tamura RN, Krischer JP, Pagnoux C, Micheletti R, Grayson PC, Chen YF, et al. A small n sequential multiple assignment randomized trial design for use in rare disease research. *Contemp Clin Trials*. 2016;46:48-51. [[FREE Full text](#)] [doi: [10.1016/j.cct.2015.11.010](https://doi.org/10.1016/j.cct.2015.11.010)] [Medline: [26586608](https://pubmed.ncbi.nlm.nih.gov/26586608/)]
75. Lu X, Nahum-Shani I, Kasari C, Lynch KG, Oslin DW, Pelham WE, et al. Comparing dynamic treatment regimes using repeated-measures outcomes: modeling considerations in SMART studies. *Stat Med*. 2016;35(10):1595-1615. [[FREE Full text](#)] [doi: [10.1002/sim.6819](https://doi.org/10.1002/sim.6819)] [Medline: [26638988](https://pubmed.ncbi.nlm.nih.gov/26638988/)]
76. Freigoun MT, Martín CA, Magann AB, Rivera DE, Phatak SS, Korinek EV, et al. System identification of just walk: a behavioral mHealth intervention for promoting physical activity. 2017 Presented at: 2017 American Control Conference (ACC); May 24-26, 2017; 116-121; Seattle, WA, USA. [doi: [10.23919/acc.2017.7962940](https://doi.org/10.23919/acc.2017.7962940)]
77. dos Santos PL, Freigoun MT, Martin CA, Rivera DE, Hekler EB, Romano RA, et al. System identification of just walk: using matchable-observable linear parametrizations. *IEEE Trans Control Syst Technol*. 2020;28(1):264-275. [[FREE Full text](#)] [doi: [10.1109/tcst.2018.2884833](https://doi.org/10.1109/tcst.2018.2884833)]
78. Hojjatinia S, Daly ER, Hnat T, Hossain SM, Kumar S, Lagoa CM, et al. Dynamic models of stress-smoking responses based on high-frequency sensor data. *NPJ Digit Med*. 2021;4(1):162. [[FREE Full text](#)] [doi: [10.1038/s41746-021-00532-2](https://doi.org/10.1038/s41746-021-00532-2)] [Medline: [34815538](https://pubmed.ncbi.nlm.nih.gov/34815538/)]
79. Chevance G, Baretta D, Heino M, Perski O, Olthof M, Klasnja P, et al. Characterizing and predicting person-specific, day-to-day, fluctuations in walking behavior. *PLoS One*. 2021;16(5):e0251659. [[FREE Full text](#)] [doi: [10.1371/journal.pone.0251659](https://doi.org/10.1371/journal.pone.0251659)] [Medline: [33989338](https://pubmed.ncbi.nlm.nih.gov/33989338/)]
80. Goldstein SP, Thomas JG, Foster GD, Turner-McGrievy G, Butryn ML, Herbert JD, et al. Refining an algorithm-powered just-in-time adaptive weight control intervention: a randomized controlled trial evaluating model performance and behavioral outcomes. *Health Informatics J*. 2020;26(4):2315-2331. [[FREE Full text](#)] [doi: [10.1177/1460458220902330](https://doi.org/10.1177/1460458220902330)] [Medline: [32026745](https://pubmed.ncbi.nlm.nih.gov/32026745/)]
81. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124. [[FREE Full text](#)] [doi: [10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124)] [Medline: [16060722](https://pubmed.ncbi.nlm.nih.gov/16060722/)]
82. Freedland KE. Progress in health-related behavioral intervention research: making it, measuring it, and meaning it. *Health Psychol*. 2022;41(1):1-12. [doi: [10.1037/hea0001160](https://doi.org/10.1037/hea0001160)] [Medline: [35113583](https://pubmed.ncbi.nlm.nih.gov/35113583/)]
83. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol*. 1998;2(2):175-220. [doi: [10.1037/1089-2680.2.2.175](https://doi.org/10.1037/1089-2680.2.2.175)]
84. Collins LM, Kugler KC. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: Advanced Topics. Cham, Switzerland. Springer International Publishing; 2018.
85. Austrian J, Mendoza F, Szerencsy A, Fenelon L, Horwitz LI, Jones S, et al. Applying A/B testing to clinical decision support: rapid randomized controlled trials. *J Med Internet Res*. 2021;23(4):e16651. [[FREE Full text](#)] [doi: [10.2196/16651](https://doi.org/10.2196/16651)] [Medline: [33835035](https://pubmed.ncbi.nlm.nih.gov/33835035/)]
86. Miller HN, Plante TB, Gleason KT, Charleston J, Mitchell CM, Miller ER, et al. A/B design testing of a clinical trial recruitment website: a pilot study to enhance the enrollment of older adults. *Contemp Clin Trials*. 2021;111:106598. [[FREE Full text](#)] [doi: [10.1016/j.cct.2021.106598](https://doi.org/10.1016/j.cct.2021.106598)] [Medline: [34653651](https://pubmed.ncbi.nlm.nih.gov/34653651/)]
87. Senderey AB, Kornitzer T, Lawrence G, Zysman H, Hallak Y, Ariely D, et al. It's how you say it: systematic A/B testing of digital messaging cut hospital no-show rates. *PLoS One*. 2020;15(6):e0234817. [[FREE Full text](#)] [doi: [10.1371/journal.pone.0234817](https://doi.org/10.1371/journal.pone.0234817)] [Medline: [32574181](https://pubmed.ncbi.nlm.nih.gov/32574181/)]
88. Rivera DE, Pew MD, Collins LM. Using engineering control principles to inform the design of adaptive interventions: a conceptual introduction. *Drug Alcohol Depend*. 2007;88(Suppl 2):S31-S40. [[FREE Full text](#)] [doi: [10.1016/j.drugalcdep.2006.10.020](https://doi.org/10.1016/j.drugalcdep.2006.10.020)] [Medline: [17169503](https://pubmed.ncbi.nlm.nih.gov/17169503/)]
89. Hekler E, Anderson CAM, Cooper LA. Is it time to restructure the National Institutes of Health? *Am J Public Health*. 2022;112(7):965-968. [[FREE Full text](#)] [doi: [10.2105/AJPH.2022.306830](https://doi.org/10.2105/AJPH.2022.306830)] [Medline: [35728037](https://pubmed.ncbi.nlm.nih.gov/35728037/)]
90. Kwasnicka D, Hoor GAT, Hekler E, Hagger MS, Kok G. Proposing a new approach to funding behavioural interventions using iterative methods. *Psychol Health*. 2021;36(7):787-791. [doi: [10.1080/08870446.2021.1945061](https://doi.org/10.1080/08870446.2021.1945061)] [Medline: [34180340](https://pubmed.ncbi.nlm.nih.gov/34180340/)]

91. Glasgow RE, Huebschmann AG, Brownson RC. Expanding the CONSORT figure: increasing transparency in reporting on external validity. *Am J Prev Med.* 2018;55(3):422-430. [doi: [10.1016/j.amepre.2018.04.044](https://doi.org/10.1016/j.amepre.2018.04.044)] [Medline: [30033029](https://pubmed.ncbi.nlm.nih.gov/30033029/)]
92. Wolfenden L, Williams CM, Wiggers J, Nathan N, Yoong SL. Improving the translation of health promotion interventions using effectiveness-implementation hybrid designs in program evaluations. *Health Promot J Austr.* 2016;27(3):204-207. [doi: [10.1071/HE16056](https://doi.org/10.1071/HE16056)] [Medline: [29241482](https://pubmed.ncbi.nlm.nih.gov/29241482/)]
93. Ullman AJ, Beidas RS, Bonafide CP. Methodological progress note: hybrid effectiveness-implementation clinical trials. *J Hosp Med.* 2022;17(11):912-916. [FREE Full text] [doi: [10.1002/jhm.12936](https://doi.org/10.1002/jhm.12936)] [Medline: [35934981](https://pubmed.ncbi.nlm.nih.gov/35934981/)]
94. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c869. [FREE Full text] [doi: [10.1136/bmj.c869](https://doi.org/10.1136/bmj.c869)] [Medline: [20332511](https://pubmed.ncbi.nlm.nih.gov/20332511/)]
95. Montgomery P, Grant S, Mayo-Wilson E, Macdonald G, Michie S, Hopewell S, et al. Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 extension. *Trials.* 2018;19(1):407. [FREE Full text] [doi: [10.1186/s13063-018-2733-1](https://doi.org/10.1186/s13063-018-2733-1)] [Medline: [30060754](https://pubmed.ncbi.nlm.nih.gov/30060754/)]
96. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for designing and evaluating feasibility pilot studies. *Med Care.* 2022;60(1):95-103. [FREE Full text] [doi: [10.1097/MLR.0000000000001664](https://doi.org/10.1097/MLR.0000000000001664)] [Medline: [34812790](https://pubmed.ncbi.nlm.nih.gov/34812790/)]
97. Pearson N, Naylor PJ, Ashe MC, Fernandez M, Yoong SL, Wolfenden L. Guidance for conducting feasibility and pilot studies for implementation trials. *Pilot Feasibility Stud.* 2020;6(1):167. [FREE Full text] [doi: [10.1186/s40814-020-00634-w](https://doi.org/10.1186/s40814-020-00634-w)] [Medline: [33292770](https://pubmed.ncbi.nlm.nih.gov/33292770/)]
98. Heino MTJ, Knittle K, Noone C, Hasselman F, Hankonen N. Studying behaviour change mechanisms under complexity. *Behav Sci (Basel).* 2021;11(5):77. [FREE Full text] [doi: [10.3390/bs11050077](https://doi.org/10.3390/bs11050077)] [Medline: [34068961](https://pubmed.ncbi.nlm.nih.gov/34068961/)]
99. Hekler EB, Klasnja P, Chevance G, Golaszewski NM, Lewis D, Sim I. Why we need a small data paradigm. *BMC Med.* 2019;17(1):133. [FREE Full text] [doi: [10.1186/s12916-019-1366-x](https://doi.org/10.1186/s12916-019-1366-x)] [Medline: [31311528](https://pubmed.ncbi.nlm.nih.gov/31311528/)]
100. Regal RE, Billi JE, Glazer HM. Phenothiazine-induced cholestatic jaundice. *Clin Pharm.* 1987;6(10):787-794. [Medline: [2905941](https://pubmed.ncbi.nlm.nih.gov/2905941/)]
101. Komatsu T. Analysis of a characteristic alteration of glycogen in the liver of alloxan diabetic rat by fasting. *Nagoya J Med Sci.* 1977;39(3-4):59-67. [Medline: [887142](https://pubmed.ncbi.nlm.nih.gov/887142/)]
102. Korinek EV, Phatak SS, Martin CA, Freigoun MT, Rivera DE, Adams MA, et al. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. *J Behav Med.* 2018;41(1):74-86. [doi: [10.1007/s10865-017-9878-3](https://doi.org/10.1007/s10865-017-9878-3)] [Medline: [28918547](https://pubmed.ncbi.nlm.nih.gov/28918547/)]
103. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. *Health Psychol.* 2015;34S:1209-1219. [FREE Full text] [doi: [10.1037/hea0000306](https://doi.org/10.1037/hea0000306)] [Medline: [26651462](https://pubmed.ncbi.nlm.nih.gov/26651462/)]
104. Palermo TM, de la Vega R, Murray C, Law E, Zhou C. A digital health psychological intervention (WebMAP mobile) for children and adolescents with chronic pain: results of a hybrid effectiveness-implementation stepped-wedge cluster randomized trial. *Pain.* 2020;161(12):2763-2774. [FREE Full text] [doi: [10.1097/j.pain.0000000000001994](https://doi.org/10.1097/j.pain.0000000000001994)] [Medline: [32658147](https://pubmed.ncbi.nlm.nih.gov/32658147/)]
105. Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. *Trials.* 2015;16:352. [FREE Full text] [doi: [10.1186/s13063-015-0842-7](https://doi.org/10.1186/s13063-015-0842-7)] [Medline: [26279154](https://pubmed.ncbi.nlm.nih.gov/26279154/)]
106. Kim M, Kim Y, Go Y, Lee S, Na M, Lee Y, et al. Multidimensional cognitive behavioral therapy for obesity applied by psychologists using a digital platform: open-label randomized controlled trial. *JMIR Mhealth Uhealth.* 2020;8(4):e14817. [FREE Full text] [doi: [10.2196/14817](https://doi.org/10.2196/14817)] [Medline: [32352391](https://pubmed.ncbi.nlm.nih.gov/32352391/)]
107. Kollins SH, DeLoss DJ, Cañadas E, Lutz J, Findling RL, Keefe RSE, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. *Lancet Digit Health.* 2020;2(4):e168-e178. [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30017-0](https://doi.org/10.1016/S2589-7500(20)30017-0)] [Medline: [33334505](https://pubmed.ncbi.nlm.nih.gov/33334505/)]
108. Luderer H, Chiodo L, Wilson A, Brezing C, Martinez S, Xiong X, et al. Patient engagement with a game-based digital therapeutic for the treatment of opioid use disorder: protocol for a randomized controlled open-label, decentralized trial. *JMIR Res Protoc.* 2022;11(1):e32759. [FREE Full text] [doi: [10.2196/32759](https://doi.org/10.2196/32759)] [Medline: [35080499](https://pubmed.ncbi.nlm.nih.gov/35080499/)]
109. Kaizer AM, Wild J, Lindsell CJ, Rice TW, Self WH, Brown S, et al. Trial of Early Antiviral Therapies during Non-hospitalized Outpatient Window (TREAT NOW) for COVID-19: a summary of the protocol and analysis plan for a decentralized randomized controlled trial. *Trials.* 2022;23(1):273. [FREE Full text] [doi: [10.1186/s13063-022-06213-z](https://doi.org/10.1186/s13063-022-06213-z)] [Medline: [35395957](https://pubmed.ncbi.nlm.nih.gov/35395957/)]
110. Fitzsimmons-Craft EE, Taylor CB, Graham AK, Sadeh-Sharvit S, Balantekin KN, Eichen DM, et al. Effectiveness of a digital cognitive behavior therapy-guided self-help intervention for eating disorders in college women: a cluster randomized clinical trial. *JAMA Netw Open.* 2020;3(8):e2015633. [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.15633](https://doi.org/10.1001/jamanetworkopen.2020.15633)] [Medline: [32865576](https://pubmed.ncbi.nlm.nih.gov/32865576/)]
111. Pallejà-Millán M, Rey-Reñones C, Uriarte MLB, Granado-Font E, Basora J, Flores-Mateo G, et al. Evaluation of the Tobstop mobile app for smoking cessation: cluster randomized controlled clinical trial. *JMIR Mhealth Uhealth.* 2020;8(6):e15951. [FREE Full text] [doi: [10.2196/15951](https://doi.org/10.2196/15951)] [Medline: [32589153](https://pubmed.ncbi.nlm.nih.gov/32589153/)]

112. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry*. 2006;59(11):990-996. [doi: [10.1016/j.biopsych.2005.09.014](https://doi.org/10.1016/j.biopsych.2005.09.014)] [Medline: [16368078](https://pubmed.ncbi.nlm.nih.gov/16368078/)]
113. Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. *Arch Gen Psychiatry*. 2006;63(5):484-489. [doi: [10.1001/archpsyc.63.5.484](https://doi.org/10.1001/archpsyc.63.5.484)] [Medline: [16651505](https://pubmed.ncbi.nlm.nih.gov/16651505/)]
114. Anvari F, Lakens D. Using anchor-based methods to determine the smallest effect size of interest. *J Exp Soc Psychol*. 2021;96:104159. [FREE Full text] [doi: [10.1016/j.jesp.2021.104159](https://doi.org/10.1016/j.jesp.2021.104159)]
115. Wang XQ, Mao SP. Comparison of morphology, pathogenicity and drug response among three isolates of schistosoma japonicum in the mainland of China. *Ann Parasitol Hum Comp*. 1989;64(2):110-119. [FREE Full text] [doi: [10.1051/parasite/1989642110](https://doi.org/10.1051/parasite/1989642110)] [Medline: [2735700](https://pubmed.ncbi.nlm.nih.gov/2735700/)]
116. 14 iconic websites that show off classic 90s web design. Webflow. 2022. URL: <https://webflow.com/blog/90s-website-design> [accessed 2024-02-02]
117. Beckman BL. What Apple, Google, and Amazon's websites looked like in 1999. Mashable. 2020. URL: <https://mashable.com/article/90s-web-design> [accessed 2024-02-02]
118. Perski O. Scientific and ethical challenges to defining what constitutes 'proportionate evidence' for the regulation and accreditation of applications to treat addiction. *Addiction*. 2021;116(12):3285-3287. [FREE Full text] [doi: [10.1111/add.15619](https://doi.org/10.1111/add.15619)] [Medline: [34231264](https://pubmed.ncbi.nlm.nih.gov/34231264/)]
119. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016. URL: <https://www.who.int/publications/i/item/9789241511766> [accessed 2024-02-02]
120. Klasnja P, Hekler EB, Korinek EV, Harlow J, Mishra SR. Toward usable evidence: optimizing knowledge accumulation in HCI research on health behavior change. *Proc SIGCHI Conf Hum Factor Comput Syst*. 2017;2017:3071-3082. [FREE Full text] [doi: [10.1145/3025453.3026013](https://doi.org/10.1145/3025453.3026013)] [Medline: [30272059](https://pubmed.ncbi.nlm.nih.gov/30272059/)]
121. Larsen KR, Michie S, Hekler EB, Gibson B, Spruijt-Metz D, Ahern D, et al. Behavior change interventions: the potential of ontologies for advancing science and practice. *J Behav Med*. 2017;40(1):6-22. [FREE Full text] [doi: [10.1007/s10865-016-9768-0](https://doi.org/10.1007/s10865-016-9768-0)] [Medline: [27481101](https://pubmed.ncbi.nlm.nih.gov/27481101/)]
122. Larsen KI, Hekler EB, Paul MJ, Gibson BS. Improving usability of social and behavioral sciences' evidence: a call to action for a national infrastructure project for mining our knowledge. *Commun Assoc Inf Syst*. 2020;46:1-17. [FREE Full text] [doi: [10.17705/1CAIS.04601](https://doi.org/10.17705/1CAIS.04601)]
123. Reardon J, Lee SSS, Goering S, Fullerton SM, Cho MK, Panofsky A, et al. Trustworthiness matters: building equitable and ethical science. *Cell*. 2023;186(5):894-898. [FREE Full text] [doi: [10.1016/j.cell.2023.01.008](https://doi.org/10.1016/j.cell.2023.01.008)] [Medline: [36724788](https://pubmed.ncbi.nlm.nih.gov/36724788/)]
124. Espie CA, Torous J, Brennan TA. Digital therapeutics should be regulated with gold-standard evidence. *Health Affairs Forefront*. 2022. URL: <https://www.healthaffairs.org/doi/10.1377/forefront.20220223.739329/> [accessed 2024-02-02]
125. Mullins CD, Abdulhalim AM, Lavalley DC. Continuous patient engagement in comparative effectiveness research. *JAMA*. 2012;307(15):1587-1588. [doi: [10.1001/jama.2012.442](https://doi.org/10.1001/jama.2012.442)] [Medline: [22511684](https://pubmed.ncbi.nlm.nih.gov/22511684/)]
126. Minkler M, Wallerstein N. *Community-Based Participatory Research for Health: From Process to Outcomes*, 2nd Edition. San Francisco, CA. Wiley; 2011.
127. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. *Acad Emerg Med*. 2015;22(6):754-756. [FREE Full text] [doi: [10.1111/acem.12692](https://doi.org/10.1111/acem.12692)] [Medline: [25997375](https://pubmed.ncbi.nlm.nih.gov/25997375/)]
128. Petersen C, Austin RR, Backonja U, Campos H, Chung AE, Hekler EB, et al. Citizen science to further precision medicine: from vision to implementation. *JAMIA Open*. 2020;3(1):2-8. [FREE Full text] [doi: [10.1093/jamiaopen/ooz060](https://doi.org/10.1093/jamiaopen/ooz060)] [Medline: [32607481](https://pubmed.ncbi.nlm.nih.gov/32607481/)]
129. Wilson BB. *Resilience for All: Striving for Equity Through Community-Driven Design*. Washington, DC. Island Press; 2018.
130. Jason L, Glantsman O, O'Brien JF, Ramian KN. *Introduction to Community Psychology: Becoming an Agent of Change*. Montreal, Quebec, Canada. Rebus Community; 2019.
131. Whittaker A. *Research Skills for Social Work*, 2nd Edition. London. Sage Publications; 2012.
132. Ellis RB, Wright J, Miller LS, Jake-Schoffman D, Hekler EB, Goldstein CM, et al. Lessons learned: beta-testing the digital health checklist for researchers prompts a call to action by behavioral scientists. *J Med Internet Res*. 2021;23(12):e25414. [FREE Full text] [doi: [10.2196/25414](https://doi.org/10.2196/25414)] [Medline: [34941548](https://pubmed.ncbi.nlm.nih.gov/34941548/)]
133. Nebeker C, Gholami M, Kareem D, Kim E. Applying a digital health checklist and readability tools to improve informed consent for digital health research. *Front Digit Health*. 2021;3:690901. [FREE Full text] [doi: [10.3389/fdgh.2021.690901](https://doi.org/10.3389/fdgh.2021.690901)] [Medline: [34713167](https://pubmed.ncbi.nlm.nih.gov/34713167/)]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials

DTx: digital therapeutics

FDA: Food and Drug Administration

MOST: Multiphase Optimization Strategy

MVP: minimal viable product
NIH: National Institutes of Health
ORBIT: Obesity-Related Behavioral Intervention Trials
Pre-Cert: Precertification
PRECIS-2: Pragmatic Explanatory Continuum Indicator Summary-2
RE-AIM: reach effectiveness, adoption, implementation, and maintenance
RWD: real-world data
RWE: real-world evidence
SMART: specific, measurable, actionable, realistic, and timely
STEEEP: safe, timely, effective, efficient, equitable, and patient-centered
WHO: World Health Organization

Edited by T Leung; submitted 21.05.23; peer-reviewed by R Barak Ventura, D Boeldt; comments to author 06.12.23; revised version received 13.01.24; accepted 29.01.24; published 05.03.24

Please cite as:

*Kim M, Patrick K, Nebeker C, Godino J, Stein S, Klasnja P, Perski O, Viglione C, Coleman A, Hekler E
The Digital Therapeutics Real-World Evidence Framework: An Approach for Guiding Evidence-Based Digital Therapeutics Design, Development, Testing, and Monitoring
J Med Internet Res 2024;26:e49208
URL: <https://www.jmir.org/2024/1/e49208>
doi: [10.2196/49208](https://doi.org/10.2196/49208)
PMID: [38441954](https://pubmed.ncbi.nlm.nih.gov/38441954/)*

©Meelim Kim, Kevin Patrick, Camille Nebeker, Job Godino, Spencer Stein, Predrag Klasnja, Olga Perski, Clare Viglione, Aaron Coleman, Eric Hekler. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 05.03.2024. 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 the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.