The Digital Therapeutics Real World Evidence Framework: An approach for guiding evidence-based DTx design, development, testing, and monitoring

Meelim Kim, Kevin Patrick, Camille Nebeker, Job Godino, Spencer Stein, Predrag Klasnja, Olga Perski, Clare Viglione, Aaron Coleman, Eric Hekler

Submitted to: Journal of Medical Internet Research
on: May 21, 2023

Table of Contents

The Digital Therapeutics Real World Evidence Framework: An approach for guiding evidence-based DTx design, development, testing, and monitoring

Meelim Kim[1] PhD; Kevin Patrick[1] MD, PhD; Camille Nebeker[1] PhD; Job Godino[2] PhD; Spencer Stein[3] MPH; Predrag Klasnja[4] PhD; Olga Perski[1] PhD; Clare Viglione[1] MPH; Aaron Coleman[5]; Eric Hekler[1] PhD

[1]Herbert Wertheim School of Public Health and Human Longevity Science University of California San Diego San Diego US
[2]Laura Rodriguez Research Institute Family Health Centers of San Diego San Diego US
[3]Spiral Health Inc. San Diego US
[4]School of Information University of Michigan Ann Arbor US
[5]Small Steps Labs LLC dba Fitabase Inc. San Diego US

Corresponding Author:
Eric Hekler PhD
Herbert Wertheim School of Public Health and Human Longevity Science
University of California San Diego
9500 Gilman Dr
MC0811
San Diego
US

Abstract

Digital Therapeutics (DTx) are seen as a promising way to provide safe, effective, accessible, sustainable, scalable, and equitable approaches to advance individual, population, and public health. Developing DTx is inherently complex in that DTx may include multiple interacting components, such as tools to support activities like medication adherence, health behavior goal-setting or self-monitoring, and algorithms that adapt provision of these according to individual needs. While myriad frameworks exist for different parts of the DTx lifecycle, to date, no single unifying framework exists to guide DTx evidence production. The purpose of this paper is to fill this gap. Specifically, we propose the DTx Real-World Evidence (RWE) Framework to provide a pragmatic, iterative, milestone-driven approach for producing RWE for DTx. While it incorporates insights from multiple fields, it uses, as its starting foundation, the Obesity-Related Behavioral Intervention Trials (ORBIT) model, but with explicit adaptations established for DTx. The DTx RWE Framework has two key elements. The first is recommendations on the use of real-world data (RWD) across the DTx lifecycle to support identifying unmet needs, establish real-world benchmarks, support on-going monitoring, and, over time, enable more rapid and resource efficient development and testing based on RWD. The second is a flowchart, which maps onto the four-phase development model as delineated by ORBIT for behavioral interventions (which was adapted from the original phases used for pharmaceuticals) but includes key adaptations relevant to RWE production for DTx. The intended audiences for this are entities that develop and market DTx and thus need RWD and community-serving organizations such as healthcare organizations and public health departments that can provide RWD. We offer the DTx RWE Framework as a unified approach that integrates best practices to evidence production for DTx, which can be used now to guide the actions of DTx companies and community-serving organizations. With that said, regulators of DTx could also consider drawing on the DTx RWE Framework to improve guidance related to DTx evidence production.

(JMIR Preprints 21/05/2023:49208)
DOI: https://doi.org/10.2196/preprints.49208

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

The Digital Therapeutics Real World Evidence Framework: An approach for guiding evidence-based DTx design, development, testing, and monitoring

*Meelim Kim, Kevin Patrick, Camille Nebeker, Job Godino, Spencer Stein, Predrag Klasnja, Olga Perski, Clare Viglione, Aaron Coleman, Eric Hekler**

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 include multiple interacting components such as tools to support activities like medication adherence, health behavior goal-setting or self-monitoring, and algorithms that adapt 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 lifecycle 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 four-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 lifecycle 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, healthcare, 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.

1. Introduction & Background

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 provision of support to each person's changing needs. Common goals for DTx include improving medication adherence or regular use of medical devices (e.g., 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 burden on healthcare 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 U.S. Food and Drug Administration (FDA) engage in to ensure the products or services in a targeted sector (e.g. 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 two overarching issues that distinguish DTx from other therapeutics commonly used in healthcare. 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 that 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 need to be interoperable with the constantly changing landscape of other software solutions used within healthcare and require high levels of software sophistication based on enterprise-grade code embedded within a robust system architecture that supports security, privacy, and on-going 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 towards trustworthiness of DTx companies, as used in the U.S. Food and Drug Administration (FDA) Pre-Certification (Pre-Cert) Program[13]; 2) utilizing 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 four-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.

2. OUR APPROACH

Based on this background, we had two primary objectives for our proposed *Digital Therapeutics Real-world Evidence (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 four-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 STEEEP targets[24] (i.e., Safe, Timely, Effective, Efficient, Equitable, and Patient-centered) as well as accessiblity[25], sustainablility[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 (e.g., Torous et al.[28,29]). 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 – which is an NIH-recognized approach for behavioral intervention development that utilizes four phases analogous to pharmaceutical development – was a foundational source. But 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. Also, the ORBIT model focuses evidence production in support of the design of a Phase III efficacy trials. This is well-matched to studying novel intervention or 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 U.S when creating the *Framework*. When necessary, the authors utilized 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 U.S. FDA Pre-Cert Program which was created to assess the credibility and readiness for a group to engage in DTx evidence production. The U.S. 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 set 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.

3. The DTx RWE Framework

The *DTx RWE Framework* (shown in Figure 1) is centered on a flowchart with four 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 specification are delineated. Phase II activities are drawn primarily from ORBIT[12] and 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 PRECIS-2[42] and RE-AIM (Reach Effectiveness, Adoption, Implementation, and Maintenance)[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 post-marketing surveillance from the World Health Organization[48]. The approach to RWD in the *Framework* draws on recent recommendations for use of RWD for pharmaceuticals[48-51] and on recommendations on open science best practices which is integrated into each phase, with additional suggested recommendations[52] summarized in the Discussion (see Section 4).

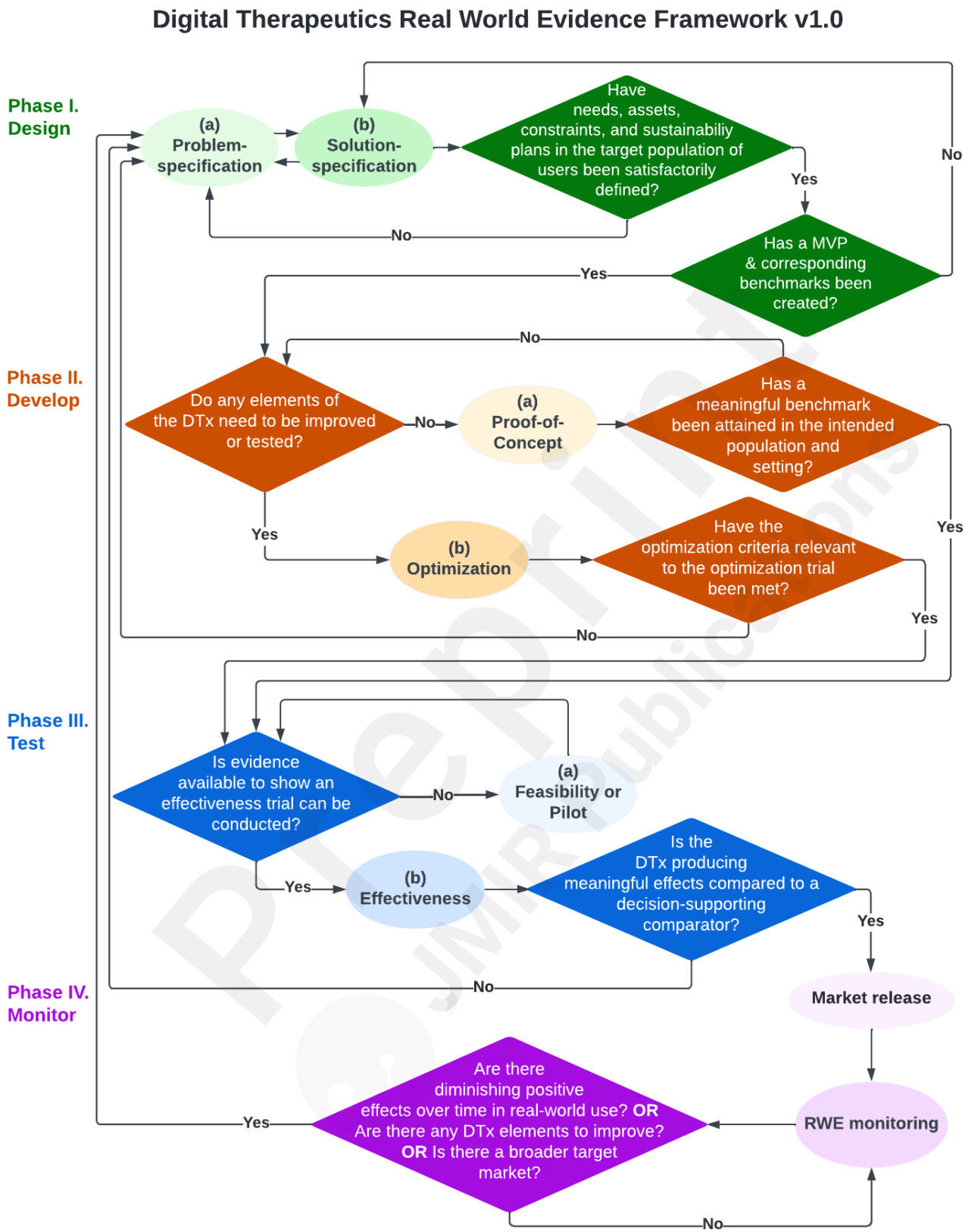


Figure 1. Digital Therapeutics Real-world Evidence Framework flowchart. *A practical, iterative, and milestone-driven guidance to producing RWE during the design, development, testing, and monitoring phases of DTx*

Oval shapes = Research/Operation Activities; Diamond shapes = Decisions/Review Activities. Acronyms: MVP = Minimal Viable Product; DTx = Digital Therapeutics; RWE = Real World Evidence.

The *Framework* has four 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 lifecycle 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 via one or more partnerships as illustrated in the *Framework* use case provided in the Online Supplement 1. With this, there are three critical roles that must be present:

1) Designate a *DTx implementor* with the capability to provide on-going, 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 (FQHC), 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 flow-chart of research activities and 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 (e.g., collection, storage, analysis, 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.

3.1 Phase I. Design

The goal of Phase I is to design the DTx product. Two activities work iteratively together: a) Problem-specification; and b) 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 (e.g., paper concepts, storyboards) to high-fidelity (fully functional code) prototypes, with this iterative work being situated within ethical practices and

exploring the potential for re-using components/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 (see Section 3.3). 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. (see a full list of plausibility targets here [note, it was labeled feasibility in this paper given that it was published prior to emerging nomenclature recommendations][56]). Progress can be determined in Phase I via specification of benchmarks that conform with the Specific, Measurable, Actionable, Realistic, and Timely (SMART) goal concept[57] but applied to establishing benchmarks. 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, setting, and answers other key questions relevant to DTx[7]. Foundational to this is the RWD from community-serving organization to identify unmet needs, current assets (e.g., standard practices, billable activities related to the targeted DTx), and constraints (e.g., likely number of billable sessions, scope of real-world need, if scaled). Beyond RWD, a mixed methods approach is recommended for this stage. This includes formative research like ethnographic studies, focus groups, and interviews; community-based participatory and community-driven methods; literature reviews, such as reviewing prior 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 via DTx that are free of “bugs” and meet basic usability requirements (e.g., good System Usability Score[59]). Ideally

benchmarks are established in relation to STEEEP targets[24] (i.e., Safe, Timely, Effective, Efficient, Equitable, and Patient-centered) as well as accessiblity[25], sustainablility[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 HbA1c among 60% of our DTx users after 6 months of intervention” for effectiveness and “less than 10% of our DTx users experienced non-serious 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/lean development practices, prototype testing and development, rapid prototype-testing, and qualitative methods[61]. See the Online Supplement 1 for an example.

3.2 Phase II. Develop

The goal of Phase II is to guide DTx development and optimization. There are two types of activities: 1) a 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 re-arranging where these trials take place. Specifically, in ORBIT, optimization trials occur late in Phase I (i.e., Phase Ib). In contrast, proof-of-concept trials occur in phase IIa in ORBIT. In the authors’ opinion, the approach used in ORBIT creates two 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 Phase 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 and, 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 towards accounting for health equity in defining a target population (e.g., targeting a population that explicitly under-utilizes 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 via 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 evidence-based selection of DTx components (factorial trials)[64-68], refinement of components, particularly those used across time (micro-randomized trials)[35,69,70], and refinement of adaptation algorithms to match provision of support to context, individual differences, and timing (micro-randomized trials[35,71], and sequential multiple assignment randomized trials[72-75]). Note also, optimization trials could also be conducted that are explicitly used to support algorithm development (e.g., 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 if 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, which tests a DTx in a small group (e.g., 10-20) in relation to a benchmark (e.g., 70% of the 10 patients shift from hypertensive to systolic blood pressure < 120mmHg). This approach is used because of two 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 suffer from the confirmation bias, which refers to the tendency for individuals to seek out or interpret evidence in ways that align with prior 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 the confirmatory bias. Quasi-experimental, single case, or within-participant designs in which the participants serve as their own controls are appropriate design options from a proof-of-concept trial. Further, mixed methods, where both qualitative and quantitative data 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, which 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, is difficult to implement in real-world contexts, is 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 include only intervention components with demonstrated effectiveness, or that the DTx can be deployed for under $50 per client, or total interactions per week with the DTx stays 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], micro-randomized 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, I. et al[40] provide guidance on when to use common optimization trial designs. See Online Supplement 1 for an example.

3.3 Phase III. Test

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: a) Feasibility or Pilot

Studies; or (b) 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 re-labeled 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, sends a message that privileges one type of evidence at the expense of other evidence. This is problematic as the evidence from the other phases are 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], via reducing confidence in any consensus statements from over-emphasis on one particularly 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 towards pragmatism via a focus on effectiveness trials, inclusion of benchmarks added on a modified CONSORT 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, use of hybrid clinical trials, which incorporate the a 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] (e.g., recruitment, eligibility, setting, etc.). 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 real-world context of DTx, we suggest a bias towards pragmaticism 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 towards 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 trials 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 (see Section 3.1). 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 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 prior to 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 is 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 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 quantify 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 (e.g., 80% of eligible clinics will take part, 80% of eligible staff will take part, 50% of eligible participants will be enrolled). As before, these benchmarks need to balance the need to be ambitious while also 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/community partner has a current standard of care and they are considering replacing it with the DTx, 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 under-powered studies to infer effect sizes[60,112,113]. Instead, what is recommended is to establish two 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. With that said, given the unreliability of small sample sizes, RWD and effect sizes from prior 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 result 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 was not a 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 Online Supplement 1 for an illustrative example.

3.4 Phase IV. Monitor

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/community partners, and to support expansion of the target market for the DTx via 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 (APIs) 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 APIs, would, at best, be perceived as "old" and, at worst, would not work. Thus, pre-specified 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 warrants 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 are meeting expectations across the STEEEP and related criteria listed earlier are critical to monitor over time. To support this, implementation science practices, particularly strategies for ongoing monitoring and thoughtful adaptation and guidance on rigorous continuous quality improvement can be gleaned from the dynamic sustainability framework[46]. Through on-going 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 re-establish a partnership with a clinical/community partner and re-engage 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 two 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 WHO, the four major components of digital health monitoring (i.e., functionality, stability, fidelity, and quality) should guide on-going monitoring, with these questions mapping on to 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 post market surveillance is being explored, and opportunities and pitfalls which are also relevant to DTx are being articulated and should be considered regarding DTx regulation[48]. Monitoring could require benchmarks be set for all key targets of evidence production (e.g., effectiveness, safety, equity, etc) as one pathway for cultivating more rigor in monitoring efforts and to reduce 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 or 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 (e.g., intervention components, adaptation algorithms) instead of the DTx package. A more detailed rationale for this is described eslewhere[7,32,47,82,120]. A second opportunity would be to link activities and efforts with on-going 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 Online Supplement 1 for an illustrative example.

4. Discussion

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 via simulated clinical trials via 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 welling institutions such the YMCA or Jewish Family Services, along with corporate wellness and related wellness programs that are not implemented by or in partnership with the healthcare 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) suggested next steps with regards to regulation and funding.

4.1 Advancing health equity through RWD

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 healthcare, 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 uses 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 DTx's and for DTx's writ 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.

4.2 Cultivating trustworthy partnerships

As presented in the section above introducing the *Framework*, we recommend a tripartite approach to its use comprised of 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* via partnerships among groups that may historically not have worked as closely together, such as industry partners working with FQHCs and supported with an academic partner, as illustrated in the Online Supplement 1.

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 been delineated already 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 determine corporate trustworthiness that were formatively tested in the U.S. FDA Pre-Cert program can be used including excellence appraisal and streamlined review elements (e.g., real-world performance plan, 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.

4.3 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 lifecycle of DTx use. Similarly, funders of DTx research and development such as the NIH, Agency for Healthcare Research and Quality (AHRQ), Patient-Centered Outcomes Research Institute (PCORI), and Health Resources and Services Administration (HRSA), could encourage applicants to use the *Framework* and if they do, then demonstrate how they propose to achieve the benchmarks that it includes.

5. 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 via DTx and we hope the *Framework* can contribute to this end.

ACKNOWLEDGEMENTS

We wish to thank Dr. Kenneth Freedland for providing expert consultation and review of earlier versions of this manuscript.

COMPETING INTERESTS

S.S. has an ownership interest in Spiral Health, Inc. A.C. has an ownership interest in Fitabase, Inc. O.P. acts as unpaid scientific advisor for the Smoke Free app. All other authors declare no competing interests.

AUTHOR CONTRIBUTIONS

M.K., K.P., C.N., and E.H. 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.

References

1 *Understanding DTx, A New Category of Medicine,*, <https://dtxalliance.org/understanding-dtx/> (2022).

2 Healthcare Brew. *Digital therapeutics get a brand new definition*, <https://www.healthcare-brew.com/stories/2023/06/12/digital-therapeutics-get-a-brand-new-definition> (2023).

3 International Organization for Standardization (ISO). *ISO/TR 11147:2023(en) Health informatics — Personalized digital health — Digital therapeutics health software systems*, <https://www.iso.org/obp/ui/#iso:std:iso:tr:11147:ed-1:v1:en> (2023).

4 Craig, P. *et al.* Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* **337**, a1655, doi:10.1136/bmj.a1655 (2008).

5 Patel, N. A. & Butte, A. J. Characteristics and challenges of the clinical pipeline of digital therapeutics. *NPJ Digit Med* **3**, 159, doi:10.1038/s41746-020-00370-8 (2020).

6 Dang, A., Arora, D. & Rane, P. Role of digital therapeutics and the changing future of healthcare. *J Family Med Prim Care* **9**, 2207-2213, doi:10.4103/jfmpc.jfmpc_105_20 (2020).

7 Murray, E. *et al.* Evaluating Digital Health Interventions: Key Questions and Approaches. *Am J Prev Med* **51**, 843-851, doi:10.1016/j.amepre.2016.06.008 (2016).

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* **19**, e232, doi:10.2196/jmir.7126 (2017).

9 Voorheis, P. *et al.* Integrating Behavioral Science and Design Thinking to Develop Mobile Health Interventions: Systematic Scoping Review. *JMIR Mhealth Uhealth* **10**, e35799, doi:10.2196/35799 (2022).

10 Torous, J. *Evaluating Apps: Making Informed Decisions with Available Real-World Data. Digital Health Interventions from Wellness to Therapeutics: Development and Dissemination.* (National Institutes of Health, Digital Health Intervention Workshop, July, 13th, 2022).

11 Miao, B. Y., Arneson, D., Wang, M. & Butte, A. J. Open challenges in developing digital therapeutics in the United States. *PLOS Digit Health* **1**, doi:10.1371/journal.pdig.0000008 (2022).

12 Czajkowski, S. M. *et al.* From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. *Health Psychol* **34**, 971-982, doi:10.1037/hea0000161 (2015).

13 U.S. Food & Drug Administration. *The Software Precertification (Pre-Cert) Pilot Program: Tailored Total Product Lifecycle Approaches and Key Findings*. September 2022)

14 Chen, Z. *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* **4**, 84, doi:10.1038/s41746-021-00452-1 (2021).

15 Swift, B. *et al.* Innovation at the Intersection of Clinical Trials and Real-World Data Science to Advance Patient Care. *Clin Transl Sci* **11**, 450-460, doi:10.1111/cts.12559 (2018).

16 Muller, P., Chandra, N. K. & Sarkar, A. Bayesian approaches to include real-world data in clinical studies. *Philos Trans A Math Phys Eng Sci* **381**, 20220158,

doi:10.1098/rsta.2022.0158 (2023).

17 Wang, C. & Rosner, G. L. A Bayesian nonparametric causal inference model for synthesizing randomized clinical trial and real-world evidence. *Stat Med* **38**, 2573-2588, doi:10.1002/sim.8134 (2019).

18 Kokol, P. Agile Software Development in Healthcare: A Synthetic Scoping Review. *Applied Sciences* **12**, 9462 (2022).

19 Van Velthoven, M. H., Smith, J., Wells, G. & Brindley, D. Digital health app development standards: a systematic review protocol. *BMJ Open* **8**, e022969, doi:10.1136/bmjopen-2018-022969 (2018).

20 Agency, P. a. M. D. *Basic principles on use of registries in approval applications*.

21 Ministry of Food and Drug Safety (MFDS). *Guideline on the Application of Real World Evidence of Medical Device*, <https://www.mfds.go.kr/brd/m_1060/view.do?seq=14328&srchFr=&srchTo=&srchWord=&srchTp[…]m_seq=0&company_cd=&company_nm=&Data_stts_gubun=C1004&page=18> (2019).

22 U.S. Food & Drug Administration. *Framework for FDA's Real-World Evidence Program*, <https://www.fda.gov/media/120060/download> (2018).

23 MobiHealthNews. *Pear Therapeutics assets sold for $6M at auction after bankruptcy*, <https://www.mobihealthnews.com/news/pear-therapeutics-assets-sold-6m-auction-after-bankruptcy> (2023).

24 Agency for Healthcare Research and Quality. *Six Domains of Healthcare Quality.*, <https://www.ahrq.gov/talkingquality/measures/six-domains.html> (2022).

25 Gulliford, M. *et al.* What does 'access to health care' mean? *J Health Serv Res Policy* **7**, 186-188, doi:10.1258/135581902760082517 (2002).

26 McCool, J. *et al.* Factors influencing the sustainability of digital health interventions in low-resource settings: Lessons from five countries. *J Glob Health* **10**, 020396, doi:10.7189/jogh.10.020396 (2020).

27 Milat, A. *et al.* Intervention Scalability Assessment Tool: A decision support tool for health policy makers and implementers. *Health Res Policy Syst* **18**, 1, doi:10.1186/s12961-019-0494-2 (2020).

28 Torous, J. B. *et al.* A Hierarchical Framework for Evaluation and Informed Decision Making Regarding Smartphone Apps for Clinical Care. *Psychiatr Serv* **69**, 498-500, doi:10.1176/appi.ps.201700423 (2018).

29 Lagan, S., Sandler, L. & Torous, J. Evaluating evaluation frameworks: a scoping review of frameworks for assessing health apps. *BMJ Open* **11**, e047001, doi:10.1136/bmjopen-2020-047001 (2021).

30 World Health Organization (WHO). *Role of policy-makers and health care leaders in implementation of the Global Patient Safety Action Plan 2021-2030: consensus statement.*, 2022).

31 National Institutes of Health (NIH). *NIH Consensus Statement*, <https://www.nichd.nih.gov/publications/pubs/pku/sub3#> (2000).

32 Patrick, K. *et al.* The Pace of Technologic Change: Implications for Digital Health Behavior Intervention Research. *Am J Prev Med* **51**, 816-824, doi:10.1016/j.amepre.2016.05.001 (2016).

33 Nebeker, C., Torous, J. & Bartlett Ellis, R. J. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Med* **17**, 137, doi:10.1186/s12916-019-1377-7 (2019).

34 Nebeker, C., Bartlett Ellis, R. J. & Torous, J. Development of a decision-making checklist tool to support technology selection in digital health research. *Transl Behav Med* **10**, 1004-1015, doi:10.1093/tbm/ibz074 (2020).

35 Klasnja, P. *et al.* Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychol* **34S**, 1220-1228, doi:10.1037/hea0000305 (2015).

36 Klasnja, P. *et al.* Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps. *Ann Behav Med* **53**, 573-582, doi:10.1093/abm/kay067 (2019).

37 Hekler, E. B. *et al.* Tutorial for Using Control Systems Engineering to Optimize Adaptive Mobile Health Interventions. *J Med Internet Res* **20**, e214, doi:10.2196/jmir.8622 (2018).

38 Design Council. *Framework for Innovation: Design Council's evolved Double Diamond*, <https://www.designcouncil.org.uk/our-work/skills-learning/tools-frameworks/framework-for-innovation-design-councils-evolved-double-diamond/> (2019).

39 Powell, L. H., Freedland K. E., Kaufmann P. G.,. *Behavioral Clinical Trials for Chronic Diseases: Scientific Foundations*. (SpringerLink, 2021).

40 Nahum-Shani, I., Dziak, J. J. & Wetter, D. W. MCMTC: A Pragmatic Framework for Selecting an Experimental Design to Inform the Development of Digital Interventions. *Front Digit Health* **4**, 798025, doi:10.3389/fdgth.2022.798025 (2022).

41 Linda M. Collins. *Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST)*. (Springer, 2018).

42 Loudon, K. *et al.* The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ : British Medical Journal* **350**, h2147, doi:10.1136/bmj.h2147 (2015).

43 Glasgow, R. E. *et al.* RE-AIM Planning and Evaluation Framework: Adapting to New Science and Practice With a 20-Year Review. *Front Public Health* **7**, 64, doi:10.3389/fpubh.2019.00064 (2019).

44 Holtrop, J. S. *et al.* Understanding and applying the RE-AIM framework: Clarifications and resources. *J Clin Transl Sci* **5**, e126, doi:10.1017/cts.2021.789 (2021).

45 Freedland, K. E. *et al.* The selection of comparators for randomized controlled trials of health-related behavioral interventions: recommendations of an NIH expert panel. *J Clin Epidemiol* **110**, 74-81, doi:10.1016/j.jclinepi.2019.02.011 (2019).

46 Chambers, D. A., Glasgow, R. E. & Stange, K. C. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. *Implement Sci* **8**, 117, doi:10.1186/1748-5908-8-117 (2013).

47 Hekler, E. B. *et al.* Agile science: creating useful products for behavior change in the real world. *Transl Behav Med* **6**, 317-328, doi:10.1007/s13142-016-0395-7 (2016).

48 Beaulieu-Jones, B. K. *et al.* Examining the Use of Real-World Evidence in the Regulatory Process. *Clin Pharmacol Ther* **107**, 843-852, doi:10.1002/cpt.1658 (2020).

49 Garrison, L. P., Jr., Neumann, P. J., Erickson, P., Marshall, D. & Mullins, C. D. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health* **10**, 326-335, doi:10.1111/j.1524-4733.2007.00186.x (2007).

50 Hekler, E., Tiro, J. A., Hunter, C. M. & Nebeker, C. Precision Health: The Role of the Social and Behavioral Sciences in Advancing the Vision. *Ann Behav Med* **54**, 805-826, doi:10.1093/abm/kaaa018 (2020).

51 Vaithinathan, A. G. & Asokan, V. Public health and precision medicine share a goal. *J Evid Based Med* **10**, 76-80, doi:10.1111/jebm.12239 (2017).

52 Burns, L. *et al.* Real-World Evidence for Regulatory Decision-Making: Guidance From Around the World. *Clin Ther* **44**, 420-437, doi:10.1016/j.clinthera.2022.01.012 (2022).

53 Chevance, G. *et al.* Digital health at the age of the Anthropocene. *Lancet Digit Health* **2**, e290-e291, doi:10.1016/S2589-7500(20)30130-8 (2020).

54 Chevance, G. *et al.* Thinking Health-related Behaviors in a Climate Change Context: A Narrative Review. *Ann Behav Med*, doi:10.1093/abm/kaac039 (2022).

55 Kwasnicka, D. *et al.* White Paper: Open Digital Health – accelerating transparent and scalable health promotion and treatment. *Health Psychology Review* **16**, 475-491, doi:10.1080/17437199.2022.2046482 (2022).

56 Bowen, D. J. *et al.* How we design feasibility studies. *Am J Prev Med* **36**, 452-457, doi:10.1016/j.amepre.2009.02.002 (2009).

57 University of California. *SMART Goals: A How to Guide*, <https://www.ucop.edu/local-human-resources/_files/performance-appraisal/How%20to%20write%20SMART%20Goals%20v2.pdf> (2016).

58 Sociocracy For All. *The Difference Between Whole-Group Consensus and Dynamic Governance/Sociocracy*, <https://www.sociocracyforall.org/the-difference-between-whole-group-consensus-and-dynamic-governance-sociocracy/> (2023).

59 usability.gov. *System Usability Scale (SUS)*, <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html> (

60 Freedland, K. E. Pilot trials in health-related behavioral intervention research: Problems, solutions, and recommendations. *Health Psychol* **39**, 851-862, doi:10.1037/hea0000946 (2020).

61 Nielsen Norman Group. *Design Thinking 101*, <https://www.nngroup.com/articles/design-thinking/> (2016).

62 Martin, C. A. *et al.* Development of a Control-Oriented Model of Social Cognitive Theory for Optimized mHealth Behavioral Interventions. *IEEE Trans Control Syst Technol* **28**, 331-346, doi:10.1109/tcst.2018.2873538 (2020).

63 Chevance, G. *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* **43**, 254-261, doi:10.1007/s10865-020-00137-2 (2020).

64 Spring, B. *et al.* A Factorial Experiment to Optimize Remotely Delivered Behavioral Treatment for Obesity: Results of the Opt-IN Study. *Obesity (Silver Spring)* **28**, 1652-1662, doi:10.1002/oby.22915 (2020).

65 Thomas, J. G. *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* **100**, 106217, doi:10.1016/j.cct.2020.106217 (2021).

66 Bennett, G. G. *et al.* Optimizing an Obesity Treatment Using the Multiphase Optimization

Strategy Framework: Protocol for a Randomized Factorial Trial. *JMIR Res Protoc* **10**, e19506, doi:10.2196/19506 (2021).

67 Cipriani, A. & Barbui, C. What is a factorial trial? *Epidemiol Psychiatr Sci* **22**, 213-215, doi:10.1017/S2045796013000231 (2013).

68 Kip, H., Da Silva, M. C., Bouman, Y. H. A., van Gemert-Pijnen, L. & Kelders, S. M. A self-control training app to increase self-control and reduce aggression - A full factorial design. *Internet Interv* **25**, 100392, doi:10.1016/j.invent.2021.100392 (2021).

69 Seewald, N. J., Smith, S. N., Lee, A. J., Klasnja, P. & Murphy, S. A. Practical Considerations for Data Collection and Management in Mobile Health Micro-randomized Trials. *Stat Biosci* **11**, 355-370, doi:10.1007/s12561-018-09228-w (2019).

70 Bell, L. *et al.* Notifications to Improve Engagement With an Alcohol Reduction App: Protocol for a Micro-Randomized Trial. *JMIR Res Protoc* **9**, e18690, doi:10.2196/18690 (2020).

71 Battalio, S. L. *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* **109**, 106534, doi:10.1016/j.cct.2021.106534 (2021).

72 Czyz, E. K. *et al.* Adaptive intervention for prevention of adolescent suicidal behavior after hospitalization: a pilot sequential multiple assignment randomized trial. *J Child Psychol Psychiatry* **62**, 1019-1031, doi:10.1111/jcpp.13383 (2021).

73 Gonze, B. B. *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* **9**, e14322, doi:10.2196/14322 (2020).

74 Tamura, R. N. *et al.* A small n sequential multiple assignment randomized trial design for use in rare disease research. *Contemp Clin Trials* **46**, 48-51, doi:10.1016/j.cct.2015.11.010 (2016).

75 Lu, X. *et al.* Comparing dynamic treatment regimes using repeated-measures outcomes: modeling considerations in SMART studies. *Stat Med* **35**, 1595-1615, doi:10.1002/sim.6819 (2016).

76 Freigoun, M. T. *et al.* System identification of Just Walk: A behavioral mHealth intervention for promoting physical activity in *2017 American Control Conference (ACC)*. 116-121.

77 Santos, P. L. d. *et al.* System Identification of Just Walk: Using Matchable-Observable Linear Parametrizations. *IEEE Transactions on Control Systems Technology* **28**, 264-275, doi:10.1109/TCST.2018.2884833 (2020).

78 Hojjatinia, S. *et al.* Dynamic models of stress-smoking responses based on high-frequency sensor data. *NPJ Digit Med* **4**, 162, doi:10.1038/s41746-021-00532-2 (2021).

79 Chevance, G. *et al.* Characterizing and predicting person-specific, day-to-day, fluctuations in walking behavior. *PLoS One* **16**, e0251659, doi:10.1371/journal.pone.0251659 (2021).

80 Goldstein, S. P. *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* **26**, 2315-2331, doi:10.1177/1460458220902330 (2020).

81 Ioannidis, J. P. Why most published research findings are false. *PLoS Med* **2**, e124, doi:10.1371/journal.pmed.0020124 (2005).

82 Freedland, K. E. Progress in health-related behavioral intervention research: Making it,

measuring it, and meaning it. *Health Psychol* **41**, 1-12, doi:10.1037/hea0001160 (2022).

83 Nickerson, R. S. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology* **2**, 175-220, doi:10.1037/1089-2680.2.2.175 (1998).

84 Linda M. Collins; Kari C. Kugler. *Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: Advanced Topics*. (Springer, 2018).

85 Austrian, J. *et al.* Applying A/B Testing to Clinical Decision Support: Rapid Randomized Controlled Trials. *J Med Internet Res* **23**, e16651, doi:10.2196/16651 (2021).

86 Miller, H. N. *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* **111**, 106598, doi:10.1016/j.cct.2021.106598 (2021).

87 Berliner Senderey, A. *et al.* It's how you say it: Systematic A/B testing of digital messaging cut hospital no-show rates. *PLoS One* **15**, e0234817, doi:10.1371/journal.pone.0234817 (2020).

88 Rivera, D. E., Pew, M. D. & Collins, L. M. Using engineering control principles to inform the design of adaptive interventions: a conceptual introduction. *Drug Alcohol Depend* **88 Suppl 2**, S31-40, doi:10.1016/j.drugalcdep.2006.10.020 (2007).

89 Hekler, E., Anderson, C. A. M. & Cooper, L. A. Is It Time to Restructure the National Institutes of Health? *Am J Public Health* **112**, 965-968, doi:10.2105/AJPH.2022.306830 (2022).

90 Kwasnicka, D., Ten Hoor, G. A., Hekler, E., Hagger, M. S. & Kok, G. Proposing a new approach to funding behavioural interventions using iterative methods. *Psychol Health* **36**, 787-791, doi:10.1080/08870446.2021.1945061 (2021).

91 Glasgow, R. E., Huebschmann, A. G. & Brownson, R. C. Expanding the CONSORT Figure: Increasing Transparency in Reporting on External Validity. *Am J Prev Med* **55**, 422-430, doi:10.1016/j.amepre.2018.04.044 (2018).

92 Wolfenden, L., Williams, C. M., Wiggers, J., Nathan, N. & Yoong, S. L. Improving the translation of health promotion interventions using effectiveness-implementation hybrid designs in program evaluations. *Health Promot J Austr* **27**, 204-207, doi:10.1071/HE16056 (2016).

93 Ullman, A. J., Beidas, R. S. & Bonafide, C. P. Methodological progress note: Hybrid effectiveness-implementation clinical trials. *J Hosp Med* **17**, 912-916, doi:10.1002/jhm.12936 (2022).

94 Moher, D. *et al.* CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, c869, doi:10.1136/bmj.c869 (2010).

95 Montgomery, P. *et al.* Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 Extension. *Trials* **19**, 407, doi:10.1186/s13063-018-2733-1 (2018).

96 Teresi, J. A., Yu, X., Stewart, A. L. & Hays, R. D. Guidelines for Designing and Evaluating Feasibility Pilot Studies. *Med Care* **60**, 95-103, doi:10.1097/MLR.0000000000001664 (2022).

97 Pearson, N. *et al.* Guidance for conducting feasibility and pilot studies for implementation trials. *Pilot Feasibility Stud* **6**, 167, doi:10.1186/s40814-020-00634-w (2020).

98 Heino, M. T. J., Knittle, K., Noone, C., Hasselman, F. & Hankonen, N. Studying Behaviour Change Mechanisms under Complexity. *Behav Sci (Basel)* **11**, doi:10.3390/bs11050077

(2021).

99 Hekler, E. B. *et al.* Why we need a small data paradigm. *BMC Med* **17**, 133, doi:10.1186/s12916-019-1366-x (2019).

100 Regal, R. E., Billi, J. E. & Glazer, H. M. Phenothiazine-induced cholestatic jaundice. *Clin Pharm* **6**, 787-794 (1987).

101 Komatsu, T. Analysis of a characteristic alteration of glycogen in the liver of alloxan diabetic rat by fasting. *Nagoya J Med Sci* **39**, 59-67 (1977).

102 Korinek, E. V. *et al.* Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. *J Behav Med* **41**, 74-86, doi:10.1007/s10865-017-9878-3 (2018).

103 Nahum-Shani, I., Hekler, E. B. & Spruijt-Metz, D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. *Health Psychol* **34S**, 1209-1219, doi:10.1037/hea0000306 (2015).

104 Palermo, T. M., 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* **161**, 2763-2774, doi:10.1097/j.pain.0000000000001994 (2020).

105 Copas, A. J. *et al.* Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. *Trials* **16**, 352, doi:10.1186/s13063-015-0842-7 (2015).

106 Kim, M. *et al.* Multidimensional Cognitive Behavioral Therapy for Obesity Applied by Psychologists Using a Digital Platform: Open-Label Randomized Controlled Trial. *JMIR Mhealth Uhealth* **8**, e14817, doi:10.2196/14817 (2020).

107 Kollins, S. H. *et al.* A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. *Lancet Digit Health* **2**, e168-e178, doi:10.1016/S2589-7500(20)30017-0 (2020).

108 Luderer, H. *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* **11**, e32759, doi:10.2196/32759 (2022).

109 Kaizer, A. M. *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* **23**, 273, doi:10.1186/s13063-022-06213-z (2022).

110 Fitzsimmons-Craft, E. E. *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* **3**, e2015633, doi:10.1001/jamanetworkopen.2020.15633 (2020).

111 Palleja-Millan, M. *et al.* Evaluation of the Tobbstop Mobile App for Smoking Cessation: Cluster Randomized Controlled Clinical Trial. *JMIR Mhealth Uhealth* **8**, e15951, doi:10.2196/15951 (2020).

112 Kraemer, H. C. & Kupfer, D. J. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry* **59**, 990-996, doi:10.1016/j.biopsych.2005.09.014 (2006).

113 Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J. & Yesavage, J. A. Caution regarding the

use of pilot studies to guide power calculations for study proposals. *Arch Gen Psychiatry* **63**, 484-489, doi:10.1001/archpsyc.63.5.484 (2006).

114 Anvari, F. & Lakens, D. Using anchor-based methods to determine the smallest effect size of interest. *Journal of Experimental Social Psychology* **96**, 104159, doi:https://doi.org/10.1016/j.jesp.2021.104159 (2021).

115 Wang, X. Q. & Mao, S. P. Comparison of morphology, pathogenicity and drug response among three isolates of Schistosoma japonicum in the mainland of China. *Ann Parasitol Hum Comp* **64**, 110-119, doi:10.1051/parasite/1989642110 (1989).

116 Webflow. *14 iconic websites that show off classic 90s web design*, <https://webflow.com/blog/90s-website-design> (2022).

117 Mashable. *What Apple, Google, and Amazon's websites looked like in 1999*, <https://mashable.com/article/90s-web-design> (2020).

118 Perski, O. Scientific and ethical challenges to defining what constitutes 'proportionate evidence' for the regulation and accreditation of applications to treat addiction. *Addiction* **116**, 3285-3287, doi:10.1111/add.15619 (2021).

119 World Helath Organization. *Monitoring and evaluating digital health interventions: A practical guide to conducting research and assessment*, <https://www.who.int/publications/i/item/9789241511766> (2016).

120 Klasnja, P., Hekler, E. B., Korinek, E. V., Harlow, J. & Mishra, S. R. Toward Usable Evidence: Optimizing Knowledge Accumulation in HCI Research on Health Behavior Change. *Proc SIGCHI Conf Hum Factor Comput Syst* **2017**, 3071-3082, doi:10.1145/3025453.3026013 (2017).

121 Larsen, K. R. *et al.* Behavior change interventions: the potential of ontologies for advancing science and practice. *J Behav Med* **40**, 6-22, doi:10.1007/s10865-016-9768-0 (2017).

122 Larsen, K., Hekler, E. B., Paul, M. J., & Gibson, B. S. Improving Usability of Social and Behavioral Sciences' Evidence: A Call to Action for a National Infrastructure Project for Mining Our Knowledge. *Communications of the Association for Information Systems* **46**, pp-pp, doi:https://doi.org/10.17705/1CAIS.04601 (2020).

123 Reardon, J. *et al.* Trustworthiness matters: Building equitable and ethical science. *Cell* **186**, 894-898, doi:10.1016/j.cell.2023.01.008 (2023).

124 Health Affairs Forefront. *Digital Therapeutics Should Be Regulated With Gold-Standard Evidence*, <https://www.healthaffairs.org/do/10.1377/forefront.20220223.739329/> (2022).

125 Mullins, C. D., Abdulhalim, A. M. & Lavallee, D. C. Continuous patient engagement in comparative effectiveness research. *JAMA* **307**, 1587-1588, doi:10.1001/jama.2012.442 (2012).

126 Minkler, M. & Wallerstein, N. *Community-Based Participatory Research for Health: From Process to Outcomes*. (Wiley, 2011).

127 Birnbaum, F., Lewis, D., Rosen, R. K. & Ranney, M. L. Patient engagement and the design of digital health. *Acad Emerg Med* **22**, 754-756, doi:10.1111/acem.12692 (2015).

128 Petersen, C. *et al.* Citizen science to further precision medicine: from vision to implementation. *JAMIA Open* **3**, 2-8, doi:10.1093/jamiaopen/ooz060 (2020).

129 Barbara Wilson. *Resilience for all: Striving for Equity Through Community-Driven Design*. (Island Press Washington, DC, 2018).

130 Jason, L., Glantsman, O., O'Brien, J. & Ramian, K. *Introduction to Community Psychology: Becoming an Agent of Change*. (2019).
131 Whittaker, A. *Research skills for social work*. (2012).
132 Bartlett Ellis, R. *et al.* Lessons Learned: Beta-Testing the Digital Health Checklist for Researchers Prompts a Call to Action by Behavioral Scientists. *J Med Internet Res* **23**, e25414, doi:10.2196/25414 (2021).
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* **3**, 690901, doi:10.3389/fdgth.2021.690901 (2021).

Online Supplement 1

This hypothetical example is based on an actual digital health company and healthcare organization, with co-authors from both (SS and JG respectively). The Digital Health start-up, Spiral Health, is seeking to become a DTx focused on providing treatment and prevention options for chronic musculoskeletal pain. Family Health Centers of San Diego (FHCSD), is the community-serving organization for the hypothetical example. FHCSD is one of the nation's ten largest Federally Qualified Health Centers (FQHCs) operating in San Diego County. Note, in this hypothetical example, we do identify a third group, which we label as "DTx Evaluators." As flagged already, this role could be filled by academics or consulting groups, or researchers could be housed at either the community-serving organization or the DTx company. In this example, the role of DTx Evaluator is being filled by an academic partner. As can be seen in the hypothetical example, we flag the Spiral Health as the DTx Implementer, the FHCSD as the Community-serving Organization, who is the key initial driver, using RWD specify where the most acute needs and opportunities for DTx exist.

The following tables describe:

- [Table 1] Partnership Formation Discussion Points
- [Table 2] Phase I. Design Example
- [Table 3] Phase II. Develop Example
- [Table 4] Phase III. Test Example
- [Table 5] Phase IV. Monitor Example

	Spiral Health	Family Health Centers of San Diego (FHCSD)
Partnership building		
Identify the Success Criteria	1) Patient-reported improvements in pain (e.g., VAS pain scale improvements) 2) Improvement in function as measured by the OLBPI 3) Increased self-efficacy in managing spine pain 4) High customer satisfaction 5) Develop a DTx that can be prescribed and reimbursed by insurance	1) Provide everyone with caring, affordable, high-quality health care and supportive services; 2) Offer financially sustainable options to uninsured, low-income, and medically underserved populations.
Boundaries of the Targeted Market	Adults with chronic musculoskeletal pain, excluding those suffering from traumatic injury or cancer	People who are medically underserved, greatly impacted by SDoH (ie FQHC-eligible patients)
	COMBINED: People who are medically underserved, impacted by SDoH (i.e., FQHC-eligible patients) and who suffer from chronic musculoskeletal pain	

Overall Leadership Structure	1) **FHCSD** leads decision-making on defining real-world success criteria, constraints, and assets. Further, FHCSD, as an FQHC, has an ethical obligation to advance the health of their patients. Thus, FHCSD can play a key role in ensuring all ethical requirements (e.g., attaining appropriate institutional review from an external party) are conducted and maintained. With that said, all parties would be responsible for following ethical practices and principles. 2) **Spiral Health** leads decision-making on DTx development and solution specification that meets FHCSD success criteria within real-world constraints and assets and aligns with the purpose of Spiral's DTx offering. 3) **DTx Evaluator,** provides expertise and guidance on the appropriate and ethical use of methods relevant to each phase. Note that the role of DTx Evaluator can shift between phases as method requirements change between phases. DTx Evaluator can be internal staff at FHCSD or Spiral or could be drawn from external partners such as academic partners or consultancy firms.

FHCSD=Family Health Center San Diego; VAS=Visual Analogues Scale; OLBPI=Oswestry Low Back Pain Index; SDoH=Social Determinants of Health; FQHC=Federally Qualified Health Center; DTx=Digital Therapeutics

Table 1. Partnership Formation Discussion Points Example

Phase I: Design	
Leadership Structure in Phase I	1) **FHCSD** is key driver of Phase I activities, supporting and organizing of patients, providers, and other stakeholders to define success, constraints, and assets. 2) **Spiral Health** engages, with the goal of specifying a solution that fits FHCSD's success criteria, constraints, and assets. 3) **DTx Evaluator**, facilitates team and partnership formation activities to facilitate robust ethical, shared decision-making and offers skills in human-centered design methods and processes.
The Type of RWD to be used	**Collect and organize relevant data from most similar standard practice to the DTx** (e.g., indicators effectiveness, attrition rates, safety risks, complications, rehospitalization rates, etc) from FHCSD's EMR, with a focus on carefully selecting a target population and setting, accounting for balancing the need to advance health equity (or at a minimum, not propagating health disparities) and the development and sustainment requirements of a DTx.
Problem Specification	**Methods used:** A series of mixed methods formative studies would be used to define problems, needs, assets, and constraints. This includes engaging in conducting data analytics on available RWD with a particular focus on identifying issues of health disparities, such as some populations not being served with current resources; needs such as some settings unable to offer in-person support; and benchmarks, such as current success rates for analogous interventions to Spiral Health among different populations. In addition, active use of qualitative methods, such as interviews, focus groups, field studies/contextual inquiry, diary studies, and direct observation, could be used among all relevant stakeholders including providers, patients, and hospital administrators, along with surveys to provide quantitative results. Finally, literature reviews, both from academic and non-academic sources, would take place to help better specify the problem and also likely needs, assets, and constraints. Using an iterative approach focused on triangulating insights from disparate sources to a share understanding, the following problems, needs, assets, and constraints are identified. **Problem:** FQHCs often lack or are limited in PT and chiropractic services compared to the need. Insurance based system often creates constraints on these services such as number of visits covered, expensive copays, and few actual minutes with a provider. Additionally, attending PT appointments can be inconvenient especially for those with transportation/time barriers. **Needs:** - A convenient, effective, and low-cost solution to manage and reduce chronic

	musculoskeletal pain that could be billable within an FQHC and fit into current PT billability options. **Assets:** - Current chronic pain management protocols. - FQHC financial models for serving FQHC-eligible patients. **Constraints:** - Patients' skepticism of the effectiveness of exercises for chronic musculoskeletal pain (negative prior experience with chronic musculoskeletal pain treatments and lack of results) - Patients having difficulty in behavioral/lifestyle changes required to commit time and energy towards consistent practice of therapeutic exercises - Different levels among patients' ability to use the technology in both assessments and following the instruction during the intervention - Clinician burnout. - Limited clinician training in managing chronic musculoskeletal pain. - Insufficient physician buy-in.
Solution Specification	**Methods used:** Mixed methods would be used, but with the methods focus more on testing the solution. RWD is used, but, increasingly, to define target populations and settings, paying particular focus to seeking to produce DTx that are serving currently underserved populations and settings in the community and clearly specified benchmarks, in terms treatment response among the target population Spiral Health. In addition, literature reviews, both from academic and non-academic sources, would take place to find currently available tools and other resources that could inform the work. Qualitative methods are used such as co-design workshops, participatory design (e.g., the formation of a diverse team with representative of all key stakeholders working together over time to design solutions that fit their real-world needs(, think aloud usability testing, A/B testing, unmoderated testing of tools, such as formative studies to examine real-world use of a Spiral Health among all relevant stakeholders including providers, patients, and hospital administrators. Using an iterative approach focused on triangulating the design of a Spiral Health that is solving the problem identified in Phase I(a) accounting for needs, assets, and constraints, while also providing more details on key issues for creating a viable sustainable solution including a sustainability plan, intervention elements of Spiral Health, and implementation strategies that could enable the Spiral Health to be used in the community-serving organization, and Specific, Measurable, Actionable, Realistic, Timely (SMART) benchmarks, which, in this example would be: **Solution:** Develop an effective, ease-of-use, affordable DTx for patients with chronic musculoskeletal pain who receive care from an FQHC. **Sustainability Plan** An approach to sustainably pay for Spiral Health services rendered within current billing constraints of an FQHC, should it be shown effective in Phase III within an FQHC population. **The intervention elements of an MVP of Spiral Health** - Success story emails - Daily text reminders - Virtual Live Coaching, as a supplement to FHCSD-delivered care - Education about the patient's body using the AI system, which helps explain the problem, what needs to change to see results, and how the exercises can make these improvements - Monthly tracking/retesting of functional movement scores to objectively show patient their progress - Clear instructional videos so the patient feels confident in performing the activities - Badges/points/rewards for daily engagement/completion of treatment plan

	Implementation elements needed to implement Spiral Health with FHCSD - Create user's guide for both patients and clinicians to enhance the easy-of-use of the tool - Find the right balance between the effectiveness of treatment and time management for treatment (e.g, provide 5 videos/day or customize the time for the patients - Add portals for clinicians to monitor patients' activities in managing their pain - Provide continuing education to the patients explaining the importance of ongoing exercise therapy - Collect the user experience feedbacks whether they would recommend to their friends/families **Benchmarks** *Effectiveness* - Over 50% of enrolled patients of Spiral Health have improvement in 2 points in VAS Pain Scale **OR** the average of 15% improvement in OLBPI after using the 16 week program of Spiral Health. *Engagement* - Over 50% of enrolled users of Spiral Health have engaged in the program at least 5 days a week over the 16 week program. *Safety* - 100% of enrolled users experience no serious adverse events linked directly to use of Spiral Health Services.
Go/No go to next phase decision:	***Question 1a:*** *Have needs, assets, constraints, and sustainability plans in the targeted population of users been satisfactorily defined?* → ***Yes:*** *Move to Question 1b* → ***No:*** *Move to Problem Specification in Phase I. Design* ***Question 1b:*** *Have benchmarks and, at minimum, an MVP version of the DTx been created?* → ***Yes:*** *Move to Phase II. Develop* → ***No:*** *Move to Solution Specification in Phase I. Design*

DTx = Digital Therapeutics; FHCSD=Family Health Center San Diego; RWD=Real-world Data; EMR=Electronic Medical Record; VAS=Visual Analogue Scale; OLBPI=Oswestry Low Back Pain Index; PT=Physical Therapy; FQHC=Federally Qualified Health Center; MVP=Minimal Viable Product

Table 2. Phase I. Design Example

Phase II: Develop	
Leadership Structure in Phase II	1) **Spiral Health** is the key driver of Phase II, in both developing and optimizing DTx to meet benchmarks. 2) **FHCSD** leads decision-making on defining benchmarks. 3) **DTx Evaluator** design and conduct a POC and/or Optimization Trial(s)
The Type of RWD to be used	**RWD** is used to specify and justify sufficient need for a **targeted population and targeted setting.** In this example, that translates to FHCSD patients eligible to work with a PT but not engaging in care. RWD also used to monitor for **unintended consequences** of use of Spiral Health, such as reduced used of other services, increases in pain reporting, hospitalizations, etc.
Decision-making Scenario 1: *(Question 2a) Do any elements of the DTx need to be improved or tested?*	SCENARIO 1: FHCSD reviews Spiral Health plans and determines that 1) if Spiral Health existed, it would be an important and valuable addition to FHCSD's care offerings **AND** 2) the benchmarks are meaningful and could plausibly be met among patients at FHCSD. Spiral Health reviews DTx elements and determines it is: 1) feasible to implement all DTx elements with fidelity and without additional need for development, **AND** 2) that they are confident that all the elements are needed. IF THE ABOVE IS TRUE THEN: **DTx Evaluator,** in partnership with Spiral

	Health and FHCSD, designs and conducts a POC at and with FHCSD.
Proof-of-concept	**DTx Evaluators** conduct a POC trial (N=10), testing if Spiral Health achieves the targeted benchmarks established in Phase I (see above), conducted at FHCSD, as an indicator of plausibility. Within POC, a mixed methods approach would be use that include gathering surveys and feedback to test assumptions on key implementation requirements such as sufficient acceptability, demand, capacity to be integrated into current work, etc. These mixed methods would incorporate feedback from all relevant stakeholders, particularly the targeted patient population, providers, and healthcare administrators. Prior to conducting the POC appropriate ethical review would take place, such as receiving approval from an Institutional Review Board.
Decision-making Scenario 2: *(Question 2a) Do any elements of the DTx need to be improved or tested?*	SCENARIO 2: FHCSD reviews Spiral Health plans and determines that: 1) if Spiral Health produced the targeted effects, it would be an important and valuable addition to their care offerings **BUT** The benchmarks are meaningful but there is insufficient evidence to conclude that Spiral Health could achieve said targets with the currently available Spiral Health DTx package. Spiral Health reviews DTx elements and determines it is: 1) plausible to implement all of DTx elements at this time **BUT** 2) Spiral Health is *NOT* confident that all the elements are needed, particularly the 'live coaching', given the added costs and impact live coaching has on scalability potential. IF THE ABOVE IS TRUE THEN: **DTx Evaluator,** in partnership with **Spiral Health** and **FHCSD,** designs and conducts a Optimization Trial at and with FHCSD.
Optimization	**DTx Evaluators** design and conduct a screening experiment/factorial as used in MOST, to produce evidence needed for guiding evidence-based optimization of the Spiral Health DTx. Within the optimization trial, a mixed methods approach would be use that include gathering surveys and feedback to test assumptions on key implementation requirements such as sufficient acceptability, demand, capacity to be integrated into current work, etc. These mixed methods would incorporate feedback from all relevant stakeholders, particularly the targeted patient population, providers, and healthcare administrators. Prior to conducting the POC appropriate ethical review would take place, such as receiving approval from an Institutional Review Board. Further, open science practices would be expected such as trial registration, open sharing of treatment protocols, and public sharing of data used for analyses, whenever possible. **Optimization criteria set by Spiral Health and FHCSD for a factorial/screening experiment as used in MOST** 1) DTx only includes components shown to improve the effectiveness of the intervention **AND** total costs remain under the plausible billable amount. 2) Is it **plausible** that benchmarks (see Phase I) would be achieved if the optimized (based on results of the screening experiment) were implemented?
Go/No go to next phase decision:	***Question 2b:*** *Has a meaningful benchmark been attained in the intended population and setting?* → ***Yes:*** *Move to Phase III. Test* → ***No:*** *Move to Question 2a* ***Question 2c****: Have the optimization criteria relevant to the optimization trial been met?* → ***Yes:*** *Move to Phase III. Test* → ***No:*** *Move to Problem Specification in Phase I. Design*

DTx=Digital Therapeutics; FHCSD=Family Health Center San Diego; RWD=Real-world Data; PT=Physical Therapy; POC=Proof-of-concept; MOST=Multiphase Optimization Strategy

Table 3. Phase II. Develop Example

Phase III: Test	
Leadership Structure in Phase III	1) **DTx Evaluators** design and conduct Feasibility/Pilot Trial and/or an Effectiveness Trial. 2) **FHCSD** decides the decision-supporting comparator and supports DTx Evaluators with conducting trial with real-world patients and access to EHR/RWD. 3) **Spiral Health** offers robust version of DTx for testing at FHCSD.
The Type of RWD to be used	RWD used to specify and justify the credibility of Spiral Health by gleaning data from 53 clinics with PT services within FHCSD and among those clinics, it is found that 80% of patients who receive care can be used to specify a benchmarks for specifying target populations and specific clinics that will be targeted to reduce health disparities.
Decision-making Scenario 1: *(Question 3a) Is evidence available to show an effectiveness trial can be conducted?*	SCENARIO 1: **DTx Evaluators,** in partnership with FHCSD and Spiral Health, review the available literature and results from the POC trial and determines there is **not yet sufficient evidence to confidently conduct a well-powered, minimal attrition, effectiveness trial at FHCSD**. IF THE ABOVE IS TRUE THEN: **DTx Evaluators**, in partnership with **FHCSD** and **Spiral Health**, design and conduct a Feasibility study, at and with FHCSD.
Trial Feasibility or Pilot* Study	- **DTx Evaluators,** in partnership with **FHCSD** and **Spiral Health**, create a protocol for a full effectiveness trial, and then **pilot tests the full effective trial protocol in a pilot study** with 40 patients at FHCSD to gather data about how feasible it is to recruit, randomize, and retain participants at FHCSD. A mixed methods approach would be use that include gathering quantitative date on key trial feasibility parameters such as recruitment rates, attrition in the two groups, and survey results to gather data on levels of usability and acceptability of the intervention, as well as qualitative results to examine issues such as fidelity of delivering both the intervention and control condition, perceived utility, and to identify and potential issues not captured with quantified data, including to qualify any quantitative results, such as insights about factors that impact recruitment rates. These mixed methods would incorporate feedback from all stakeholders, particularly the targeted patient population, providers, and healthcare administrators. Prior to conducting the study appropriate ethical review would take place, such as receiving approval from Institutional Review Board.
Decision-making Scenario 2: *(Question 3a) Is evidence available to show an effectiveness trial can be conducted?*	SCENARIO 2: **DTx Evaluators**, in partnership with **FHCSD** and **Spiral Health**, review the available literature and results from the proof of concept trial [or reviews results from a completed pilot trial, see Scenario 1] and determines there **is sufficient evidence to be confident that well-powered, minimal attrition, effectiveness trial can be conducted at FHCSD** and **that it is plausible that the intervention will produce a clinically meaningful effect that surpasses the TCS (defined via benchmark from Phase I).** **DTx Evaluators**, in partnership with FHCSD and Spiral Health,

	conduct an Effectiveness Trial.
Effectiveness	**- DTx Evaluators**, in partnership with **FHCSD** and **Spiral Health** would conduct an effectiveness trial. A mixed methods approach would be use that include gathering quantitative data on key trial feasibility parameters such as recruitment rates, attrition in the two groups, and survey results to gather data on levels of usability and acceptability of the intervention, as well as qualitative results to example issues such as fidelity of delivering both the intervention and control condition, perceived utility, and to identify any potential issues not captured with quantified data, including to qualify any quantitative results, such as insights about factors that impact recruitment rates. These mixed methods would incorporate feedback from all relevant stakeholders, particularly the targeted patient population, providers, and healthcare administrators. Prior to conducting the effectiveness trial an appropriate ethical review would take place, such as receiving approval from an Institutional Review Board. At the completion of the trial, results will be shared with all relevant stakeholders both to return information and value to all stakeholders and to invite all stakeholders to provide feedback and clarify any proposed conclusions to be advanced when the trial is reviewed for regulatory approval. Further, open science practices would be expected such as trial registration, open sharing of treatment protocols, and public sharing of data used for analyses, whenever possible. **Comparator** - Current standard of care for musculoskeletal pain is to work with a physical therapist. Within this effectiveness trial RCT, the comparison will be between standard of care PT support compared to an adapted standard of care that includes fewer PT visits (e.g., assessments, active monitoring) + Spiral Health. **Power Calculations** TCS is defined as a 15% reduction in pain measured by the OLBPI scale compared to the standard of care. A statistician determined that this pain reduction corresponds to a TCS effect size of 0.4. Based on the percentage of individuals who achieved the target reduction or better, and the effects observed in similar interventions, the results of a previous POC trial (from Phase II) suggest that this level of pain reduction is plausible. Furthermore, based on the results of a pilot study and previous trials at FHCSD, the assumption for participant attrition is set at 30%. As a result, the final target sample size is 140 participants, plus an additional 60 to account for expected attrition. Thus, the total recruitment target is 200 participants. **Benchmarks related to the modified CONSORT diagram (generalization claim tests)** - 80% of eligible staff involvement AND 80% of eligible FHCSD participants enrolled in the trial
Go/No go to next phase decision:	***Question 3b:*** *Is the DTx producing meaningful effects compared to a decision-supporting comparator?* *→ Yes: Move to Phase IV. Monitor* *→ No: Move to Problem Specification in Phase I. Design*

DTx=Digital Therapeutics; FHCSD=Family Health Center San Diego; RWD=Real-world Data; PT=Physical Therapy; TCS=Threshold of Clinical Significance; RCT=Randomized Controlled Trial

*Note, a Pilot study is a specific type of feasibility, whereby the goal is to determine if the fully powered trial can be conducted with sufficient fidelity at the setting. Other types of feasibility trials can and should be used, following emerging guidelines and best practices on this work and approach.

Table 4. Phase III. Test Example

Phase IV: Monitor	
Leadership Structure in Phase IV	1) **Spiral Health** leads deployment and scaling to populations aligned with evidence produced from Phase III. 2) **DTx Evaluators** develop a RWE monitoring plan for detecting diminishing positive effects, elements to improve within DTx, or population expansion possibilities 3) **FHCSD** becomes a customer to Spiral Health, with DTx Evaluator given access to RWD.
The Type of RWD to be used	Collect data to claim that Spiral Health **is living up to its promise in accordance with the benchmarks established in Phase I.**
Market release	Once benchmarks and clinically meaningful differences are observed, the DTx may be approved for market release to the targeted population and setting beyond FHCSD (partner).
RWE monitoring *(if any of these examples apply, return to the activities from Phase I. Design)*	**Methods used:** - Emerging data science practices for on-going monitoring would be devised and reviewed by Spiral Health, DTx Evaluators, and FHCSD. Further, open science practices would be expected such as use of data standards, best practices in data management and data informatics, open sharing of treatment protocols, and public sharing of data used for analyses, whenever possible. On-going data analytics would be run by Spiral Health, to monitor for diminishing positive effects as well as opportunities for improvement of the DTx or opportunities to broaden the target market. In addition, a plan for on-going communication (e.g., yearly reports developed by Spiral Health for FHCSD, DTx Evaluators, and relevant stakeholders including targeted end-users) would be established to support on-going accountability and also to invite communication that could inspire iterative return back to earlier phases to improve upon Spiral Health. With this, the following would be explored: **[Diminishing positive effects]** - If reduced health benefits are observed in RWD. If this happens, Spiral Health reverts to early phases with a clinical partner. Clinical partner could be FHCSD or new partner. **[The improvements in DTx elements]** - Usage results reveal spots in the use of Spiral Health that could be improved. Spiral Health identifies the possibility for new technologies to be integrated (e.g., AR/VR systems) that could better support the Spiral Health. If this happens, Spiral Health reverts to early phases with a clinical partner. Clinical partner could be FHCSD or new partner. **[A broader target market]** - Off-label uses of Spiral Health, prescribed by providers serving patients in targeted areas, suggest greater possible utility for Spiral Health. If this happens, Spiral Health reverts to early phases with a clinical partner. Clinical partner could be FHCSD or new partner.

FHCSD=Family Health Center San Diego; RWE=Real-world Evidence; RWD=Real-world Data; AR/VR=Augmented Reality/ Virtual Reality

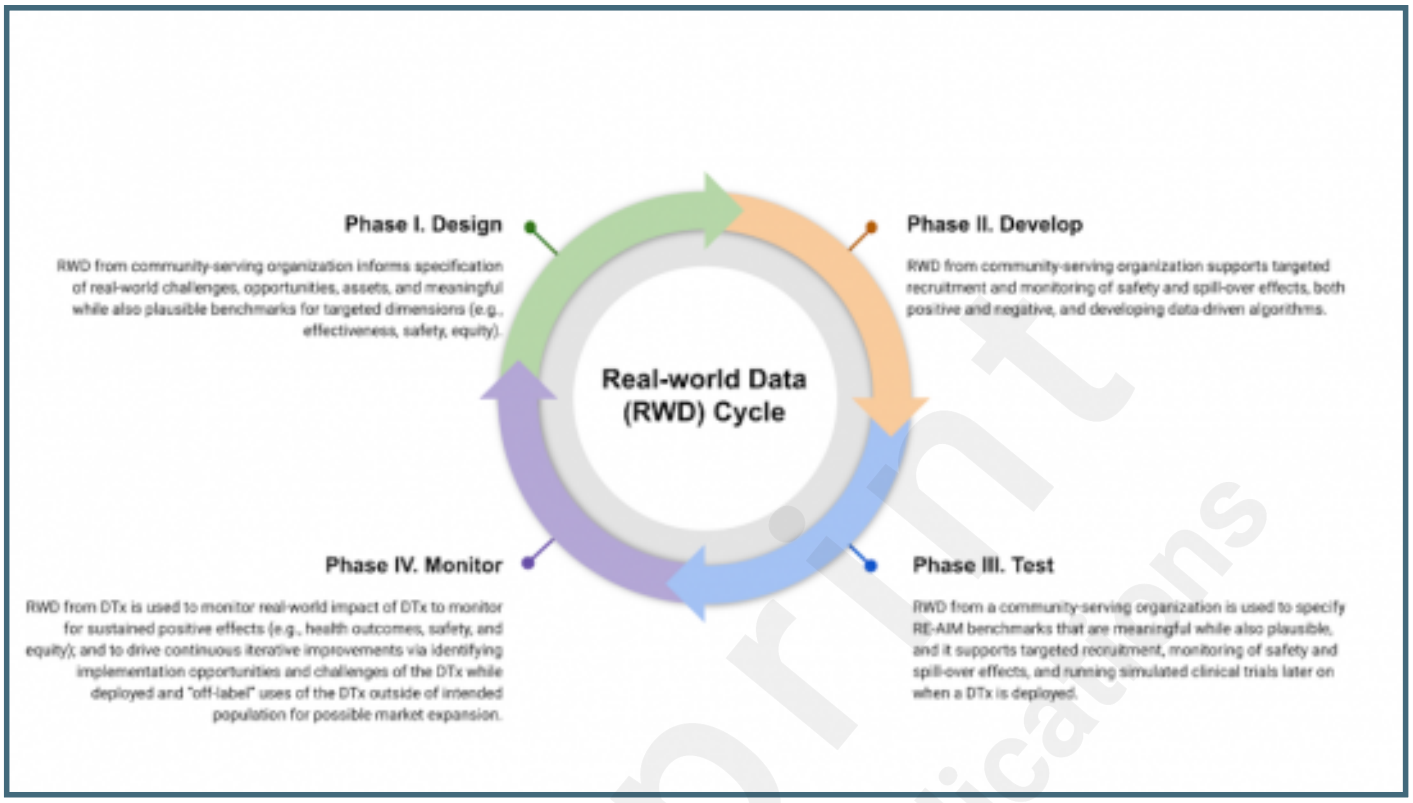
Table 5. Phase IV. Monitor Example

Supplementary Files

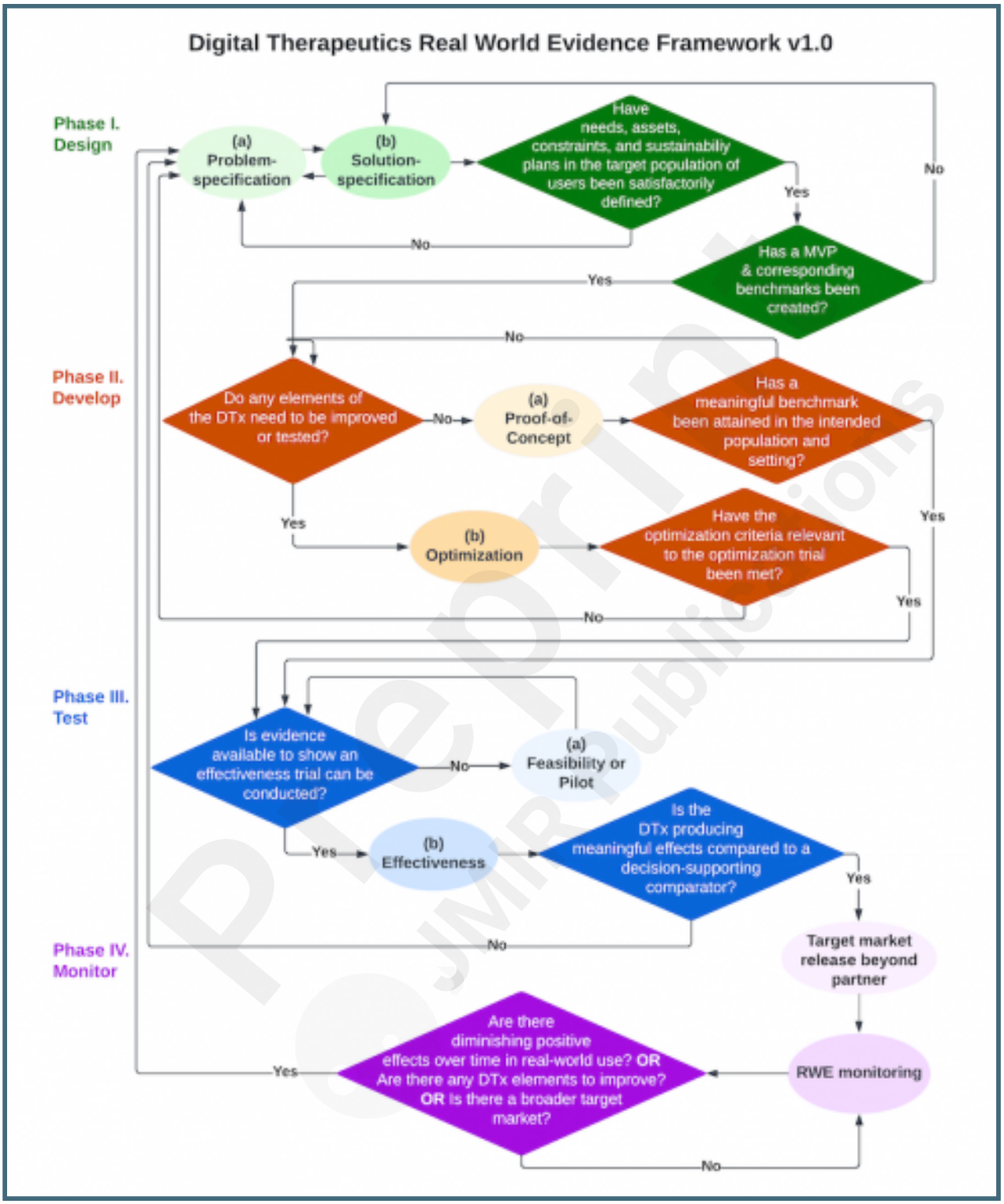
Untitled.
URL: http://asset.jmir.pub/assets/c86a00b40d13822904ed377050f8f89a.docx

Figures

Uses of RWD across the DTx lifecycle.



Digital Therapeutics Real-world Evidence Framework flowchart.



Multimedia Appendixes

The hypothetical example of DTx RWE Framework.
URL: http://asset.jmir.pub/assets/fdbca8e8fe3eae63748da45e8f301442.docx