V3+: An extension to the V3 framework to ensure user-centricity and scalability of sensor-based digital health technologies

Authors: Jessie P Bakker MS PhD (0000-0002-2976-4747)[1,2,3], Roland Barge PhD (0009-0006-9762-4769)[4], Jacob Centra PT DPT CSCS (0009-0009-6691-7923)[1], Bryan Cobb PhD FACMG (0000-0001-7147-8720)[5], Chas Cota BS (0009-0000-7215-5210)[6], Christine C Guo PhD (0000-0003-1530-0172)[7], Bert Hartog PhD (0000-0002-6631-2379)[8], Nathalie Horowicz-Mehler (0009-0005-7593-9869)[9], Elena S Izmailova PhD (0000-0002-7150-1748)[10], Nikolay V Manyakov PhD (0000-0001-9037-1400)[11], Samantha McClenahan PhD (0000-0002-1792-9700)[1], Stéphane Motola PharmD MSc (0009-0007-0416-6461)[12], Smit Patel PharmD (0000-0002-9186-0430)[1], Oana Paun MS (0009-0007-1418-5850)[13], Marian Schoone MSc (0000-0002-6604-5506)[14], Emre Sezgin PhD (0000-0001-8798-9605)[15], Thomas Switzer M.Ed (0009-0004-4998-6453)[5], Animesh Tandon MD MS (0000-0001-9769-8801)[16,17,18,19], Willem van den Brink PhD (0000-0002-4493-2382)[14], Srinivasan Vairavan PhD (0000-0001-7975-8436)[20], Benjamin Vandendriessche PhD (0000-0003-0672-0327)[1,21], Bernard Vrijens PhD (0000-0002-9090-1253)[13,22], Jennifer C Goldsack MChem MA MBA, (0000-0003-0461-0183)[1] - on behalf of the Digital Health Measurement Collaborative Community (DATAcc) hosted by the Digital Medicine Society (DiMe)

Keywords: Digital health; digital medicine; remote; decentralized; sensors; connected care; usability; ergonomics; human factors; human-centered design; user experience.

Word count: 3,567 **| Tables:** 7 **| Figures:** 5

Affiliations:

1. Digital Medicine Society, Boston MA, USA
2. Division of Sleep and Circadian Disorders, Mass General Brigham, Boston, MA, USA
3. Division of Sleep Medicine, Harvard Medical School Boston, MA, USA
4. Regeneron Pharmaceuticals Inc, New York NY, USA
5. Genentech, South San Francisco CA, USA
6. Stel Life Inc, Philadelphia PA, USA
7. Actigraph LLC, Pensacola FL, USA
8. Johnson & Johnson Innovative Medicine, Breda, the Netherlands
9. Exponent Inc, New York NY, USA
10. Koneksa Health, New York NY, USA
11. Johnson & Johnson Innovative Medicine, Beerse, Belgium
12. Sysnav Healthcare, Vernon, France
13. AARDEX Group, Liège, Belgium
14. Netherlands Organisation for Applied Scientific Research (TNO), Leiden, The Netherlands
15. Nationwide Children's Hospital, Columbus OH, USA
16. Department of Heart, Vascular, and Thoracic, Division of Cardiology and Cardiovascular Medicine, Children's Institute, Cleveland Clinic Children's, Cleveland OH, USA

17. Cleveland Clinic Children's Center for Artificial Intelligence (C4AI), Department of Heart, Vascular, and Thoracic, Children's Institute, Cleveland Clinic Children's, Cleveland OH, USA
18. Department of Pediatrics, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland OH, USA
19. Department of Biomedical Engineering, Cleveland Clinic Lerner Research Institute, Cleveland OH, USA
20. Johnson & Johnson Innovative Medicine, New Jersey NJ, USA
21. Department of Electrical, Computer, and Systems Engineering, Case Western Reserve University, Cleveland OH, USA
22. Department of Public Health, Liège University, Liège, Belgium

Corresponding author: Jessie Bakker; jessie.bakker@dimesociety.org

Author contributions: All authors contributed to the conceptual development of the V3+ framework and the drafting of this manuscript. All authors have approved the final submitted version, and agree to be accountable for all aspects of the work presented here.

Disclosures: JPB is a former employee of Philips and Signifier Medical Technologies and holds company stock or stock options, and reports consulting income from Apnimed and Koneksa Health. RB is an employee of Regeneron Pharmaceuticals Inc. and holds company stock or stock options. BC is an employee of Genentech and holds company stock. CC is an employee of Stel Life Inc. CCG is an employee of ActiGraph LLC. BH is an employee of Johnson & Johnson Innovative Medicine and holds company stock or stock options. NHM is an employee of Exponent Inc and holds company stock or stock options. ESI is an employee of Koneksa health and may own company stock. NVM is an employee of Johnson & Johnson Innovative Medicine and holds company stocks or stock options. S Matola is an employee of Sysnav Healthcare. OP is an employee of AARDEX Group. TS is an employee of Genentech and owner of shares of Roche stock. AT is a paid consultant for Synergen Health Technologies; Siemens Healthineers; Gabi Smartcare; and Medtronic (relationship is concluded). SV is an employee of Johnson & Johnson Innovative Medicine and holds company stock or stock options. B Vandendriessche is a former employee of Byteflies and holds company stock or stock options. B Vrijens is an employee of AARDEX Group and holds company stock or stock options. ES is a member of the NPJ Digital Medicine editorial board. The following authors report no disclosures: JC, JCG, S McLenahan, SP, MS, WvdB.

Acknowledgements: The authors wish to acknowledge Bethanie McCrary and Danielle Stefko for assistance with project management. On November 9th 2023, DATAcc hosted a multistakeholder workshop in order to gather and incorporate expert feedback on the content presented in this manuscript; we are grateful to the 29 individuals who participated in refining and improving our work.

Abstract

As clinical research sponsors and healthcare providers take digital clinical measures to scale, challenges related to the accessibility and usability of sensor-based digital health technologies (sDHTs) have become pressing. The heterogeneity of sDHTs - including form factors, regulatory status, and the variety of measured health concepts and applicable therapeutic areas - has led to disparate methodologies and evaluation criteria in sDHT usability studies, necessitating the need to extend the V3 framework to ensure a common lexicon and standardized approach that aligns to existing regulatory guidance and industry standards. We propose the addition of *usability validation* to the extended V3 framework, now "V3+", comprising five activities: a use specification that clearly describes the characteristics of proposed users, use environments, and the sDHT user interface; a use-related risk assessment undertaken to identify use-errors and minimize or eliminate use-related hazards; human-centered design which prioritizes the needs of users during product development; iterative formative evaluations of prototype sDHTs; and summative evaluations which aim to demonstrate sufficient usability of the sDHT within the proposed intended use or context of use. The usability validation activities presented here offer technology developers a pragmatic approach to ensuring their products can be used optimally at scale by diverse users, and provide clinical research sponsors and healthcare providers a clear roadmap for systematically identifying sDHTs that will meet the needs of their study participants and patients. Alongside the original, unchanged V3 components of verification, analytical validation, and clinical validation, implementation of the usability validation component of V3+ will ensure user-centricity and scalability of high-quality digital measurement tools, paving the way for more inclusive, reliable, and trustworthy digital measures within both clinical research and clinical care.

Introduction

Box 1: Key takeaways

1. The rapid proliferation of sensor-based digital health technologies (sDHTs) has led to some troubling examples of what may result from inadequate attention to human-centered design and usability, including compromised patient safety and extensive loss of clinical research data.
2. The latest addition to V3 - now the extended "V3+" framework - describes usability validation, a streamlined approach to evaluating the extent to which sDHTs can be used to achieve specified goals with ease, efficiency, and user satisfaction within a defined intended use or context of use.
3. We propose and describe five key activities within usability validation: development of a use specification, implementation of a use-related risk analysis, application of human-centered design including co-design with representative users, formative evaluation of prototype products, and summative evaluation of final products demonstrating sufficient usability according to pre-specified success criteria.
4. Usability validation is multidisciplinary, iterative, and cyclical in nature, and applicable to sDHTs both under development (pre-market) and commercially available (post-market).
5. We urge investigators to take a hypothesis-driven approach to usability study design, and encourage the development of standardized usability study outcome measures tailored to the unique considerations of sDHTs.
6. As was the case with the original V3 framework, V3+ is modular, meaning that changes to an sDHT limited to one component of the framework do not necessarily require collection of new evidence within the other components, optimizing efficient evidence generation supporting sDHTs as fit-for-purpose.
7. The original components of V3 - verification, analytical validation, and clinical validation - remain unchanged.
8. Implementation of the usability validation component of V3+ offers benefits to end-users and their carepartners, as well as clinicians, investigators, the technology sector, ethics committees or institutional review boards, data and safety monitoring boards, regulators, and payers, all of whom may rely on sDHT-generated data for decision-making purposes.

V3 has helped drive the increasingly broad adoption of sDHTs

Sensor-based digital health technologies (sDHTs) represent a paradigm shift in the way clinical data are captured, by providing the means to collect high-resolution data in real-world remote settings over long time-periods, and reflecting meaningful functional changes that are less prone to observer bias.[1] Within the period of 2019 - 2024, we have observed a 10-fold increase in the number of sDHT-derived measures adopted in industry-sponsored interventional trials, over 100 of which are positioned as primary endpoints.[2] In 2023, the first pivotal trial adopting a digital measure as a United States Food and Drug Administration (FDA) endorsed primary endpoint was reported,[3] and the European Medicines Agency (EMA) qualified a digital measure as a primary efficacy endpoint.[4] Although the integration of sDHTs for remote patient

monitoring and its use in clinical practice has been more gradual, implementation has been fueled by the development of a more robust infrastructure designed to tackle the complexities of large healthcare systems.[5,6]

The increasing trust and continued investment in sDHTs by healthcare providers, study sponsors, regulators, payers, and patients has been supported by V3, a modular framework for evaluating the quality of sDHTs according to:[7]

- *Verification*, which evaluates the performance of the sensor(s) against a pre-specified set of criteria;
- *Analytical validation*, which evaluates the performance of the algorithm(s) in terms of its ability to measure, detect, or predict physiological or behavioral metrics; and
- *Clinical validation*, which evaluates the extent to which the digital measure acceptably identifies, measures, or predicts a meaningful clinical, biological, physical, functional state, or experience in the specified context of use including the patient population.

Since its publication in 2020, the V3 framework has become foundational to the evaluation of sDHTs for technical, scientific, and clinical performance, having been adopted and/or referred to by individuals at the EMA,[8,9] FDA,[10] and over 250 industry and academic researchers.[11]

Development and implementation of sDHTs at scale requires an extension of V3

In the four years since the V3 framework was published, regulations focused on the use of sDHTs for remote data capture have matured[12] and reimbursement pathways have been developed, with >10 common procedural terminology codes now available for services related to digital measurement.[13] Furthermore, large-scale adoption of general wellness products has empowered individuals to monitor their own health, creating expectations for implementing sDHTs into both research and healthcare settings.[14] As clinical research sponsors and healthcare organizations take digital clinical measures to scale, challenges related to the implementation of sDHTs across diverse populations, different settings, and multifarious methodological approaches have become pressing.[15,16] For example, tremor classification data were missing for 50% of participants in the *Wearable Assessment in the Clinic and at Home in Parkinson's Disease* (WATCH-PD) study due to the inadvertent deactivation of device permissions which prevented transfer of passive data from the smartwatch to the study database.[17] Moreover, in 2023, the FDA recalled a smartphone app from an automated blood glucose monitoring and insulin delivery system because the software failed to recognize the decimal point when users entered bolus amounts <1 unit, risking the delivery of insulin doses 10 to 100 times higher than intended.[18] Although the underlying causes of these examples are multifactorial and complex, both highlight the lack of a unified set of methodological best practices to support optimal sDHT usability, defined as the extent to which an sDHT can be used to

achieve specified goals with ease, efficiency, and user satisfaction (see Table 1 for definitions).

sDHTs adopt a variety of form factors, measure a range of different health concepts, and may be regulated medical devices or non-regulated products.[19] This heterogeneity has almost certainly contributed to the disparate methodological approaches and evaluation criteria adopted in sDHT usability studies published over the last decade (systematic scoping review currently under peer-review; co-authored by AT, BRC, JC, EI, NVM, SM, SP, ES, SV, BV, and JPB). To embrace this heterogeneity and support the development of fit-for-purpose sDHTs, our goal is to build on the foundation of V3 by adding an evidence-based component to the framework addressing sDHT usability validation, as shown in Figure 1. Through dissemination of the extended V3 framework, referred to henceforth as V3+, we will ensure that sDHTs can be developed and evaluated according to state-of-the-art approaches that consider human factors engineering and human-centered design, and support the further advancement of high quality, well-validated, and easy-to-use digital measures for optimizing scientific, clinical, regulatory, and payer decision-making.

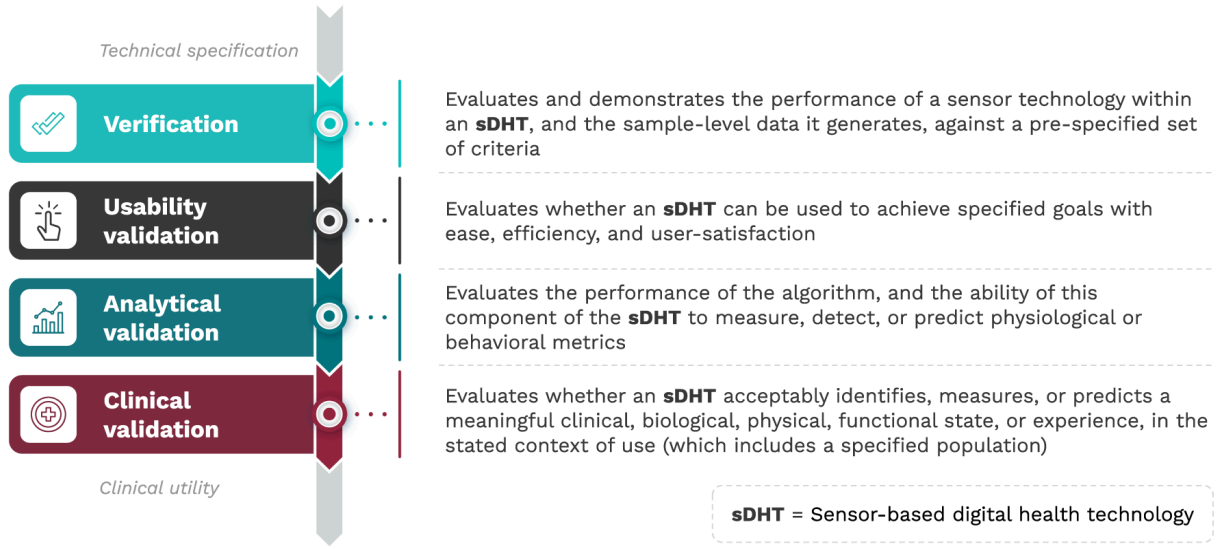

Figure 1: From V3 to V3+

The V3+ framework includes four components, with usability validation as the latest addition. Although depicted sequentially, V3+ is a modular framework as shown in Figure 5. sDHT = Sensor-based digital health technologies. See Table 1 for definitions.

Box 2: Language matters: The rationale for usability validation

Many evaluation frameworks that informed the development of V3[20–23] include **clinical utility** as the framework component directly following clinical validation: *The extent to which implementing a medical product leads to improved health outcomes or provides useful information about diagnosis, treatment, management, or prevention of disease.*[24] Notably, however, all of these frameworks position clinical utility in terms of real-world health outcomes - that is, after the point at which the product is introduced to the market - whereas V3+ applies to both pre-market and post-market sDHTs. As such, only some of the questions and processes described within clinical utility in the aforementioned frameworks, such as the importance of user-centricity and the reliance on trustworthy measurements for appropriate decision-making, are applicable to sDHTs within V3+.

We are therefore not proposing the term clinical utility to describe the fourth component of V3+; instead, we have identified **usability** as being the necessary process that allows implementation and sustainable adoption of an sDHT to be user-centric, scalable, and informative for clinical and scientific decision-making.

As is the case for analytical and clinical validation, we have adopted the term **usability validation** to refer to this entire component of the V3+ framework (see Figure 1), recognizing that several preliminary activities and studies may precede the final study/ies intended to generate evidence supporting usability.

Existing guidance and standards relevant to sDHT usability and related concepts

Recently-finalized FDA guidance clarifies that an sDHT used for remote data acquisition in a clinical investigation may be a regulated medical device per the definition under Section 201(h) of the FD&C Act, or a non-regulated product such as a general wellness product.[19] Many, but not all, of the latter are developed for and marketed primarily towards consumers, and are generally considered to be low-risk products.[25] In addition, sDHTs that are intended for use as medical devices may be considered high risk (such as some implantables[26]) or low/moderate risk (which may be the case for some ingestibles,[27] wearables,[28] and ambient[29] tools, depending on the nature of the product itself and the developer's claims).

Given this substantial variability in the ways sDHTs are regulated and marketed, best practices relevant to sDHT usability and related concepts may be found in numerous global regulatory documents[19,30-37] as well as industry standards developed for medical devices (IEC 62366-1:2015[38] and IEC/TR 62366-2:2016[39]), interactive systems (ISO 9241-210:2019)[40], everyday products (ISO 20282-1:2006[41]), consumer products (ISO/TS 20282-2:2013[42]), and others.[43] Although the recommendations in these documents are not completely aligned, many elements of the usability validation component of V3+ described below have been drawn from these existing descriptions of best practices and we have endeavored to remain consistent with the concepts and terminology wherever possible.

Box 3: What technologies does V3+ apply to?

The V3+ framework applies to **sensor-based digital health technologies (sDHTs)**, defined as *connected digital medicine products that process data captured by mobile sensors using algorithms to generate measures of behavioral and/or physiological function*.[7] In the original V3 paper, these technologies were referred to as biometric monitoring technologies (BioMeTs).

sDHTs may be:

- **Wearable:** Worn on the body, such as a chest strap or adhesive patch;
- **Implantable:** Implanted into the body, such as a loop recorder;
- **Ingestible:** Swallowed and excreted, such as a core body temperature sensor; *or*
- **Ambient:** Placed in the environment and including both passive and active data capture, such as a microphone, a connected pharmaceutical package or drug delivery device, a camera, a mattress pad, or connected weight scales.

sDHTs include **hardware**, **firmware**, and **software** components. These components may be combined into a single product, or split across multiple products; for example, by using the hardware/firmware of one sDHT to collect signals for analysis in a separate sDHT software application.

V3+ applies to *all* sDHTs used for generating digital clinical measures, regardless of a tool's regulatory status or the way(s) in which it is marketed, purchased, or accessed.

The usability validation component of V3+ is comprised of five key activities

Following the development of a proposed intended use statement, which is applicable to all four components of V3+ and which we recommend developing for both regulated and non-regulated sDHTs, the usability validation process includes five key activities as shown in Figure 2.

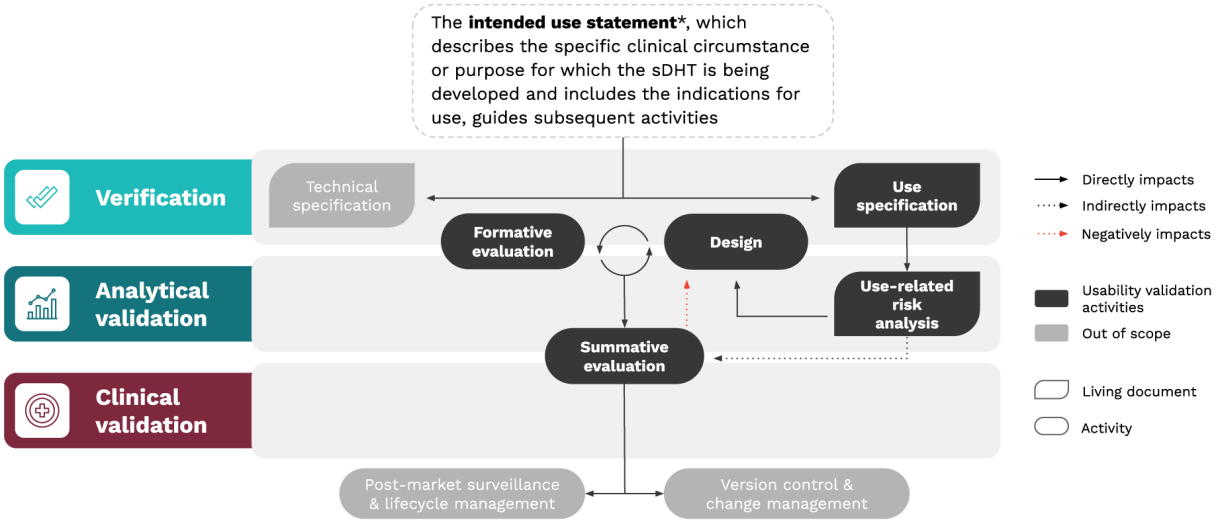

Figure 2: The five key activities of the usability validation component of V3+

This figure depicts the five usability validation activities of V3+ in the order in which they are typically undertaken, as well as the approximate alignment with verification, analytical validation, and clinical validation is also depicted. Importantly, usability validation is rarely linear, and therefore the timing and order of activities will vary case-by-case. See Table 1 for definitions.

** The intended use statement is a key component of the labeling of regulated medical devices. We recommend development of an equivalent statement for non-regulated sDHTs.*

Key activity 1: Develop the use specification

The intended use statement has a direct impact on both the technical specification (a comprehensive description of the sDHT, including but not limited to the dimensions and materials; the units of the sampled data, sampling frequency, and the sampling range of each sensor; battery life; data storage; data transmission protocols; environmental limits; and other technical information) and the use specification (a living document containing a comprehensive description of who the intended sDHT user groups are; where, when, and how each user group will interact with the sDHT including the data generated; and their motivations for doing so (may also be defined using 'use cases'). The process for creating the use specification, the first activity of usability validation, is summarized in Table 2.

Key activity 2: Conduct a use-related risk analysis

Along with the initial sDHT design, the use specification contains the information necessary to complete a preliminary version of the use-related risk analysis, which is undertaken iteratively and collaboratively to identify foreseeable risks associated with

the use of the sDHT and develop a plan for minimizing or eliminating known risks. The analysis should address reasonably foreseeable misuse, including use of the sDHT by unintended users, distinct from abnormal use which involves intentional reckless use or sabotage.[30,34,35]

The analysis begins by compiling a list of all user tasks and potential use-errors, which are actions (or lack thereof) which may result in a use-related hazard (see Table 1 for definitions, and note that "use-error" is preferable to "user-error" as it avoids the implication that the user is at fault). Once identified, use-related hazards should be categorized according to the seriousness of the potential harm in order to identify critical tasks[30] (see Figure 3). Established risk management approaches also account for the severity of potential harm and the likelihood of occurrence.[44] Optimal sDHT design involves minimizing or eliminating all possible use-errors; however, this is particularly important for critical tasks. As described in FDA guidance,[30] the ideal approach is for use-errors to be 'designed out', known as inherent safety by design; however, if this is not feasible, protective measures such as alarms or error messages may be adopted. The least favorable measures include the provision of instructions to avoid use-errors which can be included in the user manual or user training.

Importantly, harm may arise not only as a result of use-errors, but as a result of poor design leading to sub-optimal sDHT adherence and consequently, excessive missing data. Harm resulting from missing or unreliable sDHT data may include false-negative diagnostic tests, missed signs of clinical deterioration requiring intervention, or inappropriate treatment-related decisions.[45] Adoption of human-centered design principles, described below, will ensure prioritization of user satisfaction, thereby supporting user engagement and data quality.

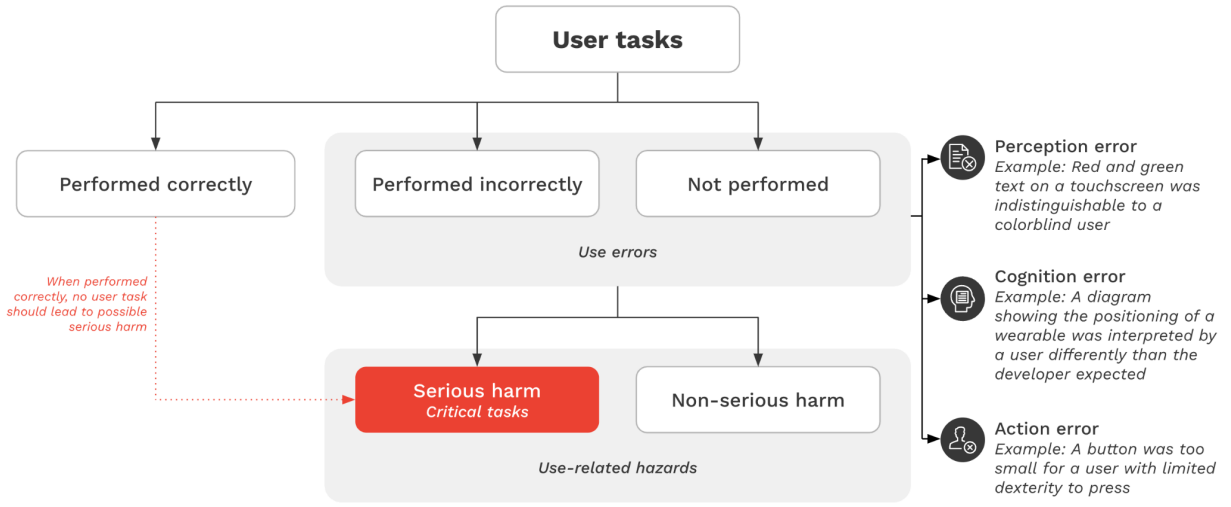

Figure 3: Relationships between the elements of a use-related risk analysis

A use-error is an action or lack of action which may result in a use-related hazard. An error that does not lead to foreseeable harm does not need to be described in the use-related risk analysis; however, such errors should be accounted for during the design process to optimize usability. See Table 1 for further definitions.

The three examples illustrated in this figure relate to the end-users of sDHTs; however, use-errors apply to all sDHT users including carepartners, clinicians, researchers, and administrators.

Key activity 3: Apply human-centered design

Optimizing sDHT usability involves more than just minimizing use-errors and use-related hazards; the goal is to create tools that are tailored to the user's needs, resulting in products that are intuitive, accessible, and enjoyable to use for a sustained period within the user population of interest. Adopting human-centered design, which prioritizes the needs, capabilities, and behaviors of users during the design process,[40] can result in products with superior usability, typically associated with greater accessibility, engagement and adherence.[46,47]

As described in Table 3, this approach involves understanding the user perspective, designing for their needs, and collaborating with representative users as partners throughout the entire end-to-end design process. This is particularly important for sDHTs, where tension may exist between the technical specification and user requirements. Thus, human-centered design considerations should be adopted throughout the entire sDHT development process, with successive prototypes assessed during formative evaluations, described below.

Key activity 4: Conduct iterative formative evaluation of sDHT prototypes

Formative evaluations include any activity or research study undertaken with the goals of describing user tasks, identifying use-errors, and gathering the information necessary to inform design improvements.[30,34,35] The circular arrow in Figure 2 indicates that formative evaluation and sDHT hardware, software, and workflow design proceed iteratively as prototypes become increasingly mature. Similarly, formative evaluations will allow the use-related risk analysis to be updated as new use-errors are discovered, which can then be addressed in an updated design and assessed during a subsequent formative evaluation. This iterative process continues until the sDHT demonstrates sufficient usability to progress to summative evaluation, described below. Common data capture methods for formative evaluations are described in Table 4.

Formative evaluation is typically conducted as a series of incremental steps, beginning with small samples (for example, $n \leq 5$) before recruiting larger samples of participants that increasingly represent the diversity of the intended end-user population, as well as all other user groups described in the use specification (see Table 2). We recommend that sDHT remote data capture be undertaken as soon as is feasible during the formative evaluation stage, in order to address any use-errors or technical difficulties associated with data storage and transfer while the sDHT design is flexible enough to accommodate required modifications.

Key activity 5: Complete summative evaluation of the final sDHT

The purpose of summative evaluations (referred to by the FDA as human factors validation studies[30]) is to demonstrate that the final (or production-equivalent[48]) version of the sDHT - including all components, accessories, packaging, instructions for use, additional documentation, and user training - is sufficiently usable within the proposed intended use including the intended user population(s).[34,35]

Summative studies should be designed to evaluate all essential user tasks including critical tasks, and all components of the user interface. Recommended methods include simulated-use and actual-use, during which the study participants interact with the sDHT independently and naturally without prompts or feedback from study staff (see Table 4).[30,34,35] The setting and duration of use should ideally reflect the conditions of actual use, and it is critical that the study design includes remote data capture without expert supervision beyond the training and support that is intended to be offered to intended users, such as helpdesk troubleshooting.

As is the case with analytical validation, the summative study sample(s) should represent the intended use population; however, a unique characteristic of usability validation studies is that all user groups identified in the use specification - such as carepartners, clinicians, investigators, research staff, and administrative staff - must be studied (see Table 2).[30,34,35] Within each user group, efforts should be made to

ensure socio-demographic diversity, as well as diversity across the user characteristics that the investigators believe will have the greatest impact on sDHT use. For example, body habitus might be important for ensuring appropriate sizing of a wearable sDHT, while dexterity, which can be impacted by age or certain health conditions, might be important for an sDHT that requires a user to input data on a small touchscreen.

Summative evaluation is the final activity in the usability validation component of V3+, and as is the case with verification and analytical validation, quantitative pass/fail criteria should be specified *a priori*. Table 5 provides real-world examples where each of the five usability validation activities have been applied during development or evaluation of sDHTs.

Usability validation is applicable to pre- and post-market sDHTs

As is the case for the original V3 framework, there are two primary use cases to which V3+ usability validation is applicable. The first relates to technologies that are under development (that is, pre-market), which may be completely novel products or next-generation versions of existing products. In this scenario, the technology developer is the party responsible for undertaking the processes described in V3+. The second use case is for technologies that are commercially available (that is, post-market). In this scenario, the party interested in deploying the sDHT - either the clinical research sponsor or the healthcare provider - is responsible for ensuring that the product has been assessed satisfactorily according to V3+. In addition, it is recommended that where possible sponsors and healthcare providers establish a collaborative relationship with the sDHT developer so that next-generation design changes can be implemented to improve usability and ensure that previously unidentified use-errors are accounted for.

Figure 2 and Figure 4 depict the elements of the usability validation component of V3+ that are applicable to pre-market and post-market sDHTs, respectively. The activities are identical in each case; the difference is that only a subset are applicable to post-market sDHTs, as there is no immediate opportunity to modify the product design or undertake formative evaluation. In addition, there are scenarios which may be considered a hybrid between pre- and post-market application of usability validation. For example, the developer of a pre-market stand-alone sDHT software application that relies on analyzing signals captured from a post-market hardware-based sDHT may proceed through the steps shown in Figure 2 including iterative design and formative evaluation of the software, but they will have no opportunity to modify the technical specification of the hardware.

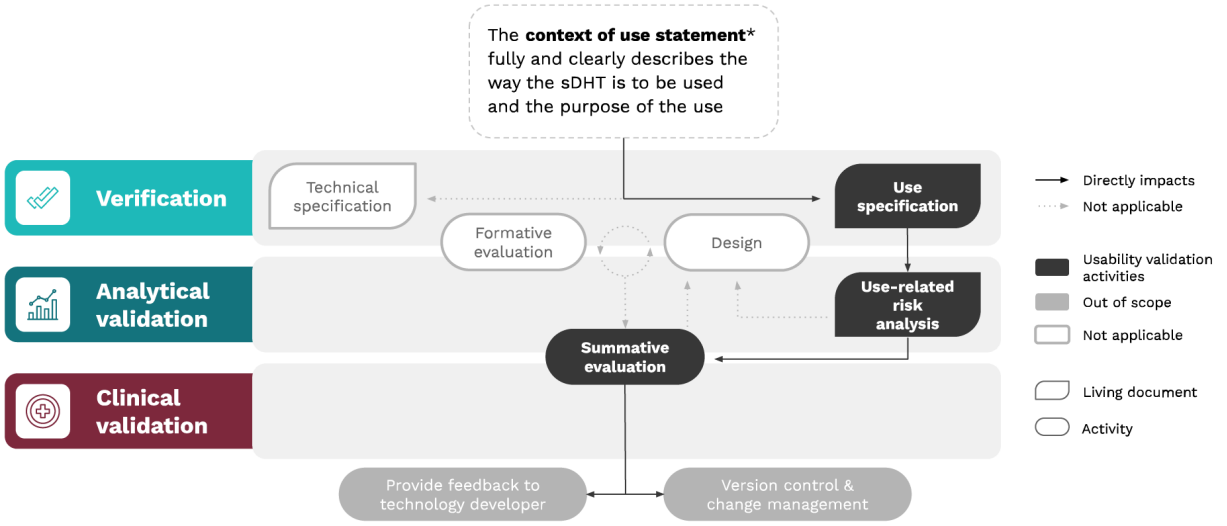

Figure 4: The three V3+ usability validation activities applicable to post-market sDHTs

This figure depicts the subset of V3+ usability validation activities that are applicable to post-market sDHTs, during which there is no immediate opportunity to modify the product design or undertake formative evaluation. See Table 1 for definitions.

** The context of use statement should be compared against the original intended use (or equivalent statement) of the sDHT; this gap analysis will guide subsequent activities.*

The activities undertaken during usability validation of a post-market sDHT are driven by a gap analysis comparing the original intended use statement and the current context of use statement. The gaps between how the sDHT was developed and the way that it is planned to be used will guide the responsible stakeholder as they determine whether existing usability data are sufficient to demonstrate fit-for-purpose, or whether further evidence is required. If the latter, the use specification and use-related risk analysis should be undertaken as described above (see Table 2 and Figure 3), before proceeding directly to summative evaluation. In many cases, summative evaluation may take the form of a bridging study designed to gather validation evidence addressing only the gaps between the intended use and context of use, thereby building on the existing evidence base.

The context of use is also important in terms of implementation in research versus healthcare settings. For example, users have expressed that the ability to access, understand, and potentially act on their own sDHT-derived clinical data is a major driver of user satisfaction and engagement;[49] however, real-time data sharing may not be suitable during a research study due to the risk of introducing behavior change.[37] As a result, perceived usefulness and user satisfaction may differ substantially between settings.

Table 6 summarizes the process of usability validation, while Table 7 describes the application of usability validation in practice.

Box 4: The sDHT intended use and context of use

The V3+ process for *pre-market* sDHTs begins with defining the **intended use statement**, which describes the specific clinical circumstances or purpose for which the sDHT is being developed and includes the indications for use. For implementation of *post-market* sDHTs, the V3+ process begins with defining the **context of use statement**, which fully and clearly describes the way the sDHT is to be used and the purpose of the use.

Intended use:

- *What* does the sDHT do?
- *Who* are the intended users?
- *Where* should the sDHT be used?
- *When* should the sDHT be used?
- *How* should the sDHT be used?

Context of use:

- *What* will the sDHT be used for?
- *Who* are the population(s) of interest?
- *Where* will the sDHT be used?
- *When* will the sDHT be used?
- *How* will the sDHT be used?

The intended use and context of use statements address similar questions; the difference is that the former summarizes the technology developer's claim/s or statement/s regarding what the sDHT does, whereas the latter describes the manner in which the sDHT will be implemented by a clinical study sponsor or healthcare provider.

'Intended use' and 'context of use' are regulatory terms which we recognize are not used with respect to non-regulated products; however, the questions addressed in each statement are critical for all four components of V3+. We therefore recommend that equivalent statements be developed for general wellness or other products for which regulations are either not applicable or not enforced.

Evaluation criteria, sample size, and statistical considerations

Several models for evaluating usability and related concepts are available, such as the Technology Acceptance Model[50] and the Unified Theory of Acceptance and Use of Technology,[51] and it is up to each investigative team to determine the most appropriate evaluation criteria for their formative and summative studies. In addition to ease of use, efficiency, and user satisfaction, common usability evaluation criteria include learnability, memorability, usefulness, use-errors, and ease of error recovery (defined in Table 1). In some cases it may be possible to leverage the sDHT backend infrastructure to capture additional metrics, such as the time taken to complete specific tasks.

While the sample size of formative evaluations is typically small, and primarily driven by the extent to which new use-errors are uncovered, the sample size of summative studies depends on the study objectives and the nature of the analyses. Regardless,

all user groups described in the use specification should be specified during protocol development, and in many cases it will be appropriate to recruit subgroups within each user group; for example, based on language, educational background, or disease severity.

Two influential papers published in the early 1990s[52,53] led to the 'rule of thumb' still in use today that 80% of use-errors can be uncovered by assessing 5 - 10 participants per user group, depending on the likelihood of problem detection. While minimization of use-errors is important, sample sizes of this magnitude do not allow for sufficient diversity to understand generalizability of study results. We therefore encourage investigators to take a hypothesis-driven approach to summative usability validation studies where appropriate, requiring a power calculation based on preliminary data demonstrating a meaningful effect size along with a robust statistical analysis plan. The sample size and recruitment plan developed for a research protocol should account for representation and diversity in addition to confidence levels and error margins. If a summative study is not hypothesis-driven, descriptive statistics are likely more appropriate than statistical testing.

Integration and alignment of V3+ with the original V3 framework

An important characteristic of the V3 framework - which also applies to V3+ - is that it is modular, meaning that changes limited to one component do not necessarily require collection of new evidence within the other components. As shown in Figure 5, which is an expanded version of Figure 3 in Goldsack *et al.*,[7] changing the use specification is a prompt to either repeat usability validation, or compile documentation demonstrating that existing usability data is generalizable to the latest use specification. Expansion to a new population, however, may prompt the need to repeat usability, analytical, and clinical validation, and in many cases it may be possible to incorporate multiple objectives into a single study.

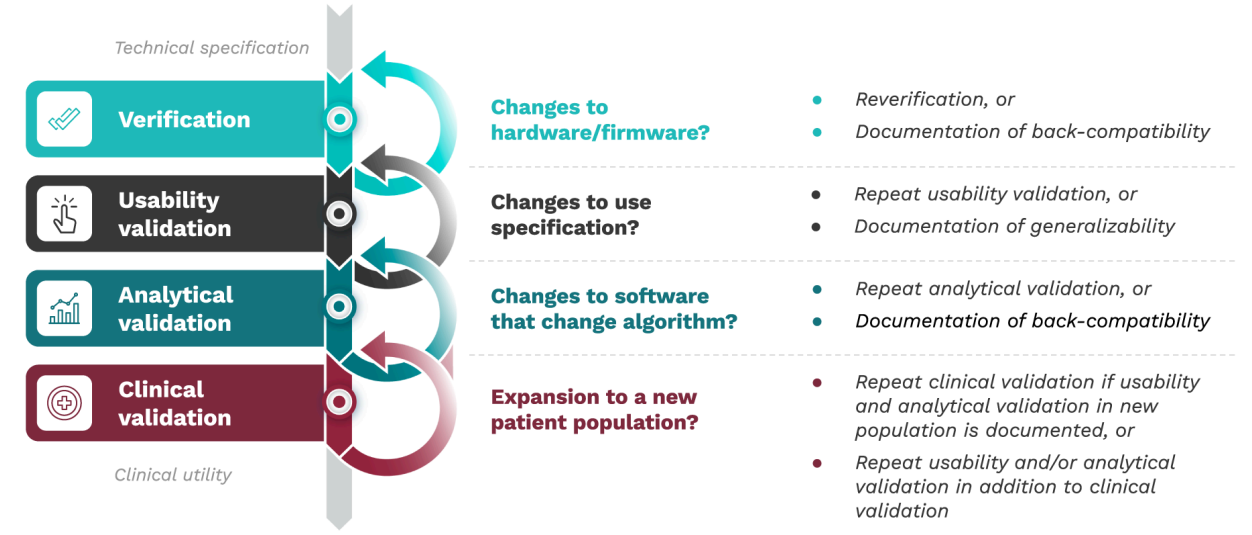

Figure 5: V3+ in action

This figure is an expanded version of Figure 3 in Goldsack et al.,[7] emphasizing the modular nature of V3+; that is, changes made within one component of the framework will require repeat testing or documentation of generalizability of that component, but may not always require repetition of the full evidence stack.

Conclusions and future directions

The implementation of sDHTs into both clinical research and healthcare has expanded rapidly in recent years, creating the need for a common lexicon and standardized approach to human-centered design and usability validation that is aligned to global regulatory guidance and industry standards. In addition to ensuring accessibility and user-centricity of sDHTs, and that appropriate scientific and clinical decisions are made on the basis of sDHT-derived data, following a pragmatic approach to usability validation offers many advantages to research study sponsors and clinicians; for example, improved workflow efficiency, reduced risk of missing data, improved clinical decision-making, and cost reduction resulting from the avoidance of unnecessary product re-design or inappropriate product selection.

The usability validation component of V3+ emphasizes the importance of identifying quantitative pre-specified pass/fail criteria, particularly during the design of summative evaluations, which can be challenging to define given the relative lack of well-established and widely-adopted sDHT-specific usability study outcomes. We therefore encourage the development and psychometric evaluation of standardized usability measures that address the unique considerations of sDHTs, including prolonged use in unsupervised environments and industry benchmarking, which will allow direct comparison both across tools and over time. Such measures can only

become widely adopted if usability studies are made available in the public domain, ideally in the peer-reviewed literature, and we therefore encourage sDHT developers and evaluators to publish their usability research including negative studies. Similarly, DATAcc by DiMe is committed to expanding our database of case studies as an open-access resource to allow the digital medicine community to share experiences, best practices, and lessons learned.[54]

Although V3+ has been developed specifically for sDHTs, many of the principles and activities described in the framework are applicable to other products and workflows. As the field of digital medicine continues to mature, we anticipate the need to apply similarly streamlined approaches in emerging areas such as AI/ML applications in healthcare, and encourage other professional bodies to build on our work. For example, there is scope to explore the design process in digital health in more depth.[47] Our hope is that by developing recommendations for usability validation best practices and supporting their implementation, sDHT user experience will become a key product differentiator as it has for consumer electronics. When all stakeholders demand optimal sDHT usability as a critical requirement rather than a nice-to-have, investment in human-centered design and robust usability validation will follow, ensuring optimal care for the diverse patient populations we serve.

Table 1: Glossary of terms.

Abnormal use: Intentional reckless use or sabotage,[30,34,35] beyond the scope of the *use-related risk analysis*.

Actionability: The extent to which *users* of diverse backgrounds, languages, and varying levels of health literacy understand the actions or *user tasks* they should complete in response to clinical data or other information presented to them,[55] typically assessed in a *knowledge task study*.

Clinical utility: The extent to which implementing an sDHT leads to improved health outcomes or provides useful information about diagnosis, treatment, management, or prevention of a disease.[24]

Cognitive walkthrough: A *formative evaluation* in which A) *usability* experts break down *user tasks* and identify possible *use-errors*[34] and/or B) usability experts guide *users* through user tasks while encouraging users to think aloud.[30,35]

Context of use: A statement that fully and clearly describes the way the sDHT is to be used and the purpose of the use.[56]

Contextual inquiry: Observation of *users* interacting with a functional sDHT in the intended *use environment*, with staff asking questions during or after use.[30,34]

Critical task: A *user task* that, if not performed or performed incorrectly, would or could lead to serious harm.[30,34,35]

Ease of use: The *ease* with which a *user* is able to perform *user tasks*,[57] captured through self-report (such as the mental demand or effort required to complete a task) or objective measures (such as the number of actions, number of attempts, or time required to complete a task).

Efficiency: The *ease* with which a *user* is able to perform *user tasks* after having learned how to use the sDHT.[57]

End-user: A user from whom sDHT-derived clinical data are captured; that is, the patient or participant.

Error recovery: The ability of a *user* to make a correction following a *use-error* in order to complete a *user task*.[57]

Fit-for-purpose: A conclusion that the level of validation associated with an sDHT is sufficient to support its *context of use*.[56]

Formative evaluation: A research study or activity undertaken to evaluate *usability* of a prototype sDHT, with the goals of understanding *user* interactions with the sDHT and gathering information to inform design modifications.[30,34,35]

Gap analysis: A systematic approach to comparing two or more statements or scenarios. For the *usability validation* component of V3+, a gap analysis will identify differences between the *intended use* and *context of use* statements.

Human factors: The application of knowledge about human behavior, abilities, limitations, and other characteristics of *users* to the design and development of an sDHT to optimize *usability* within a defined *intended use* or *context of use*. This definition incorporates terminology and concepts from FDA,[58] MHRA,[34] and NMPA (translated).[35]

Human-centered design: An approach to interactive systems that aims to make systems usable and useful by focusing on the *users*, their needs and requirements, and by applying *human factors* and *usability* knowledge and techniques.[40]

Indications for use: A statement that describes the disease or condition the sDHT is designed to diagnose, treat, prevent, cure, or mitigate, including a description of the patient or participant population for which the sDHT is intended.[59]

Intended use: A statement that describes the specific clinical circumstance or purpose for which an sDHT is being developed, including the *indications for use*.[60]

Knowledge task study: A study undertaken to assess *understandability* and *actionability*.[30]

Learnability: The *ease* with which a *user* is able to perform *user tasks* during their first encounter with the sDHT.[57]

Memorability: The *ease* with which a *user* is able to perform *user tasks* after a period of non-use, assessed in a test-retest paradigm.[57]

Production-equivalent: A sample sDHT of final design assembled in a way that differs from, but is equivalent to, the manufacturing processes used for the marketed sDHT.[61]

Sensor-based digital health technologies: Connected digital medicine products that process data captured by mobile sensors using algorithms to generate measures of behavioral and/or physiological function, also referred to as biometric monitoring technology.[7]

Summative evaluation: A research study undertaken on a *production-equivalent* or marketed sDHT including all components of the *user interface*, with the goal of demonstrating *usability* amongst a representative sample under conditions reflecting the *intended use* or *context of use,*[34,35] referred to by the FDA as 'human factors validation.[30]

Technical specification: A comprehensive description of the sDHT dimensions and materials; the units of the sampled data, sampling frequency, and the sampling range of each sensor; battery life; data storage; data transmission protocols; environmental limits; and other technical information.

Understandability: The extent to which *users* of diverse backgrounds, languages, and varying levels of health literacy understand the clinical data or other information (such as instructions, cautions, warnings, and contraindications) presented to them,[55] typically assessed in a *knowledge task study*.

Usability: The extent to which an sDHT can be used to achieve specified goals with *ease*, *efficiency*, and *user satisfaction* within a defined *intended use* or *context of use*. This definition incorporates terminology and concepts from FDA,[58] MHRA,[34] NMPA (translated),[35] and ISO 9241-210:2019.[40]

Usability validation: Evaluation and demonstration of *usability*.

Use environment: The setting(s) in which the sDHT is intended to be used.[30,34,35]

Use-error: An action or lack of action which may result in a *use-related hazard*.[30] "Use-error" is preferable to "user-error" as it avoids the implication that the user is at fault.

Use-related hazard: A source of potential harm resulting from a *use-error*. Use-related hazards are those associated with *user* interactions, rather than issues associated with sDHT technical performance or hazards such as sharp edges, unsafe operating temperatures, or non-biocompatible materials.[30]

Use-related risk analysis: A living document describing reasonably foreseeable risks associated with use of an sDHT, and a detailed plan to mitigate those risks.[34,62]

Use scenario: Rich descriptions of several likely *use environments* and how interactions with the sDHT might differ between them.[30,34,35]

Use specification: A living document containing detailed descriptions of all *user* groups, all *use environments*, and all aspects of the sDHT *user interface*.[34]

Usefulness: The extent to which a *user* finds the sDHT, or its specific features/functions, to be valuable, productive, and/or helpful.[63]

User: Any individual who may interact with an sDHT as part of normal use, including the *end-user* and their care-partner(s) as well as individuals acting in a professional capacity such as those in clinical, research, and/or administrative roles.[30,34,35]

User interface: All points of interaction a *user* may have with the sDHT as a holistic system, including hardware, software, all components and accessories, packaging, instructions for use and other documentation, and user training.[30,34,35]

User satisfaction: The extent to which a *user* finds the sDHT to be pleasant to use,[57] which may reflect trust, comfort, aesthetics, engagement, desirability, emotional response/s, and other considerations.

User task: An action or set of actions performed by a *user* to achieve a specific goal, often referred to simply as 'task'.[30]

Definitions presented in this glossary describe each term as used in V3+, and are not always verbatim from the referenced source.

Table 2: Creating the use specification

Part One: Identify all user groups and compile a series of representative user descriptions

End-users are the individuals from whom sDHT-derived clinical data will be captured. In a research study the end-users are the participants including healthy controls, while in clinical settings the end-users are the patients being cared for. Additional sDHT users might include carepartners, clinicians, investigators, research staff, and administrative staff.

For each user group, describe aspects including but not limited to socio-demographics and cultural customs; health and technology literacy; physical, sensory, and cognitive capabilities; anthropometry; disease characteristics including comorbidities; and vulnerable populations. Not all characteristics will apply to all user groups.

Part Two: Identify all likely use environments and compile a series of use scenarios

sDHTs may be used in many environments such as the home, clinic/lab settings, assisted living facilities, workplaces, educational institutions, community and leisure spaces, and during transit.

For each use environment, describe aspects including but not limited to the temperature, humidity, lighting, noise levels, space availability, cleanliness and sterility, privacy and the presence of other individuals, security and risk of theft, power and network availability, and the presence of (and interoperability with) other sDHTs.

Part Three: Compile an in-depth description of all aspects of the sDHT user interface

List the various ways users might interact with the sDHT (including hardware, software, and all accessories and components such as packaging, chargers and cables, batteries, replacement/consumable parts) and their motivations for doing so. Consider visual, auditory, and tactile cues. Compile all written materials and instructions, and describe what training, troubleshooting, and support will be available to each user group (typically identified and documented as part of a training needs analysis).

Finally, use the information from all three parts of the use specification to describe the various foreseeable interactions that users from within each user group will have with the sDHT.

Table 3: Principles of human-centered design

Human-centered design describes an iterative approach to inclusive sDHT design and development undertaken by a multidisciplinary team including user-representatives to optimize usability within a defined intended use or context of use. As described in ISO 9241-210:2019,[40] optimal human-centered design is:

- **Empathetic**: Understand the user's needs, behaviors, emotions, and capabilities.
- **Holistic**: Consider the system and the end-to-end user journey throughout the design process.
- **Iterative**: Undertake a circular loop of designing, prototyping, testing, and refining based on user feedback.
- **User-centric**: Improve usability by capturing user feedback in real-world settings, noting that 'users' includes all individuals who may interact with the sDHT including end-users (patients/participants), carepartners, clinicians, researchers, and administrators.
- **Inclusive**: Collaborate with individuals representing all user groups and elicit their feedback throughout;
- **Multidisciplinary**: Collaborate with colleagues from various disciplines to develop innovative solutions.

Table 4: A comparison of formative and summative evaluations

	Formative evaluations	***Summative evaluations***
Purpose	Describe user tasks and identify use-errors Inform iterative design modifications	Demonstrate usability amongst a representative sample under conditions reflecting the intended use or context of use
Activities	Activities such as consumer preference testing or market research Institutional Review Board or ethics committee approved or exempt research studies	
sDHT design	Prototypes with no, partial, or full functionality	Final or production-equivalent version Marketed version
Example Procedures	Cognitive walkthrough:[a] Study staff guide each participant through user tasks while encouraging users to think aloud *or* usability experts break down user tasks and identify possible use-errors Contextual inquiry: Study staff observe users interacting with the sDHT in the intended use environment, asking questions during and/or after use Knowledge task studies: Assessment of users' understandability and actionability Simulated-use, closely mimicking conditions reflecting the intended use or context of use, with no involvement of study staff Actual-use with no involvement of study staff	
Example usability testing methods	Interviews Focus groups Qualitative and quantitative surveys Real-time verbalization (think-aloud) Observation and/or timing of user tasks, use-errors, and recovery from use-errors Captured by the sDHT itself; for example, logging use-errors or timing user tasks. Product-related considerations such as connectivity loss, app crashes, and page load times may also be informative.	
Sample	Typically begin with small samples, becoming larger and increasingly similar to intended users as iterative evaluations continue.	Representative of all user groups described in the use specification.
Setting	Typically begin in-lab before progressing to remote data capture.	Representative of, or generalizable to, all use environments described in the use specification.

Duration	Typically begin with single visits before progressing to longer evaluations.	Ideally reflecting the period of intended use, including extended durations.
Evaluation criteria	Typically focused on assessing use-errors rather than formal evaluation criteria.	Quantitative pass/fail criteria must be pre-specified.

[a] *Global regulatory guidance documents define cognitive walkthrough differently; however, both approaches are suitable procedures for formative evaluations. See Table 1 for additional definitions.*

Table 5: Real-world case studies in which usability validation activities have been applied

Usability validation activities	***Case studies***
Key activity 1: Use specification	A 2018 publication by Pillalamarri *et al.*[64] provides an overview of the design approach taken during development of the Omnipod DASH Insulet Management System (Insulet Corp, Billerica MA, USA). Table 1 provides an abridged summary of representative end-users, including their characteristics, needs, and motivations for using the sDHT.
Key activity 2: Use-related risk analysis	A 2017 publication by Preusse *et al.*[65] describes a heuristic analysis, in which one expert evaluated the design of a commercially-available sDHT (Fitbit One; Fitbit Inc, San Francisco CA, USA) against Nielsen's usability heuristics,[66] followed by a secondary review undertaken by two additional experts. Violations were presented in Table 1 along with examples and possible use-errors that may result.
Key activity 3: Human-centered design	A 2018 publication by Standoli *et al.*[67] describes the process of co-designing a novel wearable sDHT developed as part of a weight-loss intervention for adolescents. A total of 407 representative end-users contributed to the sDHT design by expressing their preferences for the form factor and providing sketches. The process was conducted over multiple phases which allowed the developers to incorporate feedback iteratively as prototypes were developed.
Key activity 4; Formative evaluation	A 2020 publication by Stubberud *et al.*[68] describes the development of an sDHT smartphone app conducted over three iterative formative evaluation cycles. Usability data were captured using contextual inquiry during the first two phases, followed by a 2-week period of actual use in the intended use environment. Improvements were made to the app between each phase by incorporating user feedback.
Key activity 5: Summative evaluation	A 2022 publication by Domingos *et al.*[69] describes a summative evaluation of the Mi Band 2 (Xiaomi Inc, Beijing, China), in which 110 older adults used the sDHT in the intended use environment for 15 days before completing three validated usability surveys (Technology Acceptance Model; System Usability Scale; User Satisfaction Evaluation Questionnaire). Study hypotheses were clearly described, a power calculation was performed *a priori*, and statistical data were presented addressing each hypothesis.

Table 6: Summary of usability validation

Who?	Human factors engineers, user experience experts, designers, hardware/software developers, data scientists/analysts/statisticians, clinical researchers, and user representatives.
What?	Study protocol for data capture from human participants including the pre-specified pass/fail criteria, or procedural description of an activity other than a formal research study such as a consumer preference test or market research survey. Study report or other documentation containing sufficient information to determine the descriptive characteristics of user groups, circumstances and setting(s) of data capture, evaluation criteria, and descriptive statistics or statistical testing as appropriate.
When?	Iteratively throughout sDHT prototype development, and following finalization of the sDHT design.
Where?	Research or clinical laboratories and remote data capture environments.
Why?	To evaluate the extent to which an sDHT can be used to achieve specified goals with ease, efficiency, and user satisfaction.

Table 7: Usability validation in practice

Documentation you can expect	Documentation of activities or studies should include: • Use specification; see Table 2. • Use-related risk analysis; see Figure 3. • Summative study protocol(s) • Summative study IRB/ethics committee documentation • Study report • White paper • Peer-reviewed manuscript • Regulatory submission, if applicable.
Questions answered by usability validation	Can the sDHT be used to achieve specified goals with ease, efficiency, and user satisfaction, within the stated intended use or context of use which includes a description of all intended users?

References

1. Goldsack, J. C., Dowling, A. V., Samuelson, D., Patrick-Lake, B. & Clay, I. Evaluation, acceptance, and qualification of digital measures: from proof of concept to endpoint. *Digital biomarkers* **5**, (2021).
2. Taguibao, C. Celebrating the fourth anniversary of the Library of Digital Endpoints! *Digital Medicine Society (DiMe)* https://dimesociety.org/celebrating-the-fourth-anniversary-of-the-library-of-digital-endpoints/ (2023).
3. Bellerophon. A study to assess pulsed inhaled nitric oxide in subjects with pulmonary fibrosis at risk for pulmonary hypertension (REBUILD). *Clinical Trials* https://clinicaltrials.gov/study/NCT03267108 (2023).
4. Servais, L. *et al.* First regulatory qualification of a digital primary endpoint to measure treatment efficacy in DMD. *Nat. Med.* **29**, (2023).
5. Li, R. C., Asch, S. M. & Shah, N. H. Developing a delivery science for artificial intelligence in healthcare. *npj Digit. Med.* **3**, (2020).
6. Marwaha, J. S., Landman, A. B., Brat, G. A., Dunn, T. & Gordon, W. J. Deploying digital health tools within large, complex health systems: key considerations for adoption and implementation. *npj Digit. Med.* **5**, (2022).
7. Goldsack, J. C. *et al.* Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *npj Digit. Med.* **3**, (2020).
8. Mantua, V., Arango, C., Balabanov, P. & Butlen-Ducuing, F. Digital health technologies in clinical trials for central nervous system drugs: an EU regulatory perspective. *Nat. Rev. Drug Discov.* **20**, (2021).
9. Colloud, S. *et al.* Evolving regulatory perspectives on digital health technologies for medicinal product development. *npj Digit. Med.* **6**, (2023).
10. Sacks, L. & Kunkoski, E. Digital health technology to measure drug efficacy in clinical trials for Parkinson's disease: a regulatory perspective. *J. Parkinsons. Dis.* **11**, (2021).
11. Digital Medicine Society (DiMe). Resources in action. *Digital Medicine Society (DiMe)* https://dimesociety.org/resources-in-action/ (2022).
12. Center for Drug Evaluation & Research. Digital health technologies (DHTs) for drug development. *U.S. Food and Drug Administration* https://www.fda.gov/science-research/science-and-research-special-topics/digital-health-technologies-dhts-drug-development (2023).
13. American Medical Association. Future of health- commercial payer coverage for digital medicine codes. *American Medical Association* https://www.ama-assn.org/system/files/issue-brief-commercial-payer-coverage-digital-care.pdf (2023).
14. Kang, H. S. & Exworthy, M. Wearing the future-wearables to empower users to take greater responsibility for their health and care: scoping review. *JMIR mHealth*

and uHealth **10**, (2022).

15. Adedinsewo, D. *et al.* Health disparities, clinical trials, and the digital divide. *Mayo Clin. Proc.* **98**, (2023).
16. Landers, M., Dorsey, R. & Saria, S. Digital endpoints: definition, benefits, and current barriers in accelerating development and adoption. *Digital biomarkers* **5**, (2021).
17. Adams, J. L. *et al.* Using a smartwatch and smartphone to assess early Parkinson's disease in the WATCH-PD study. *NPJ Parkinson's disease* **9**, (2023).
18. U.S Food and Drug Administration. Class 1 device recall Omnipod 5 app. *U.S. Food and Drug Administration* https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfRES/res.cfm?id=204607 (2024).
19. U.S. Food and Drug Administration. Digital health technologies for remote data acquisition in clinical investigations- guidance for industry, investigators, and other stakeholders. *U.S. Food and Drug Administration* https://www.fda.gov/media/155022/download (2023).
20. Fryback, D. G. & Thornbury, J. R. The efficacy of diagnostic imaging. *Med. Decis. Making* **11**, (1991).
21. Haddow, J. E., Palomaki, G. E., Haddow, J. & Palomaki, G. E. ACCE: a model process for evaluating data on emerging genetic tests. https://www.scienceopen.com/document?vid=397ab49d-0b83-45cc-9ad0-c45b20c833e6 (2003).
22. National Academies of Sciences, Engineering, and Medicine. An evidence framework for genetic testing. *The National Academies Press*. 10.17226/24632 (2017).
23. IVD Clinical Evidence Working Group. Framework for developing clinical evidence for regulatory and coverage assessments in in vitro diagnostics (IVDs). *MDIC* https://mdic.org/resource/framework-for-developing-clinical-evidence-for-regulatory-and-coverage-assessments-in-in-vitro-diagnostics-ivds/ (2019).
24. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) resource. *National Library of Medicine* https://www.ncbi.nlm.nih.gov/books/NBK326791/ (U.S. Food and Drug Administration and National Institutes of Health, 2016).
25. U.S. Food and Drug Administration. General wellness: policy for low risk devices-guidance for industry and FDA staff. *U.S. Food and Drug Administration* https://www.fda.gov/media/90652/download (2019).
26. U.S Food and Drug Administration. Premarket approval (PMA). *U.S Food and Drug Administration* https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P160048S026 (2023).
27. Center for Devices & Radiological Health. Ingestible telemetric gastrointestinal capsule imaging system - final class II special controls guidance document for

industry and FDA. *U.S. Food and Drug Administration* https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/ingestible-telemetric-gastrointestinal-capsule-imaging-system-final-class-ii-special-controls (2021).

28. U.S. Food and Drug Administration. Device classification under section 513(f)(2)(De Novo). *U.S. Food and Drug Administration* https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?id=DEN200011 (2023).
29. U.S. Food and Drug Administration. 510(k) premarket notification. *U.S. Food and Drug Administration* https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K150557 (2015).
30. U.S Food and Drug Administration. Applying human factors and usability engineering to medical devices- guidance for industry and FDA staff. *U.S. Food and Drug Administration* https://www.fda.gov/media/80481/download (2016).
31. U.S Food and Drug Administration. Human factors studies and related clinical study considerations in combination product design and development- draft guidance for industry and FDA staff. *U.S Food and Drug Administration* https://www.fda.gov/media/96018/download (2016).
32. U.S Food and Drug Administration. Comparative analyses and related comparative use human factors studies for a drug-device combination product submitted in an ANDA: draft guidance for industry. *U.S Food and Drug Administration* https://www.fda.gov/media/102349/download (2017).
33. European Medicines Agency. Quality documentation for medicinal products when used with a medical device - scientific guideline. *European Medicines Agency* https://www.ema.europa.eu/en/quality-documentation-medicinal-products-when-used-medical-device-scientific-guideline (2022).
34. Department of Health and Social Care. Guidance on applying human factors and usability engineering to medical devices including drug-device combination products in Great Britain. *Medicines, Healthcare products Regulatory Agency* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/970563/Human-Factors_Medical-Devices_v2.0.pdf (2021).
35. National Medical Products Association. Guiding principles for technical review of human factors design of medical devices (DiMe, Trans.). *Digital Medicine Society (DiMe)* https://datacc.dimesociety.org/wp-content/uploads/2023/09/NMPA-Human-Factors-Guidance-English-Translation-FINAL.pdf (accessed Feb 21, 2024).
36. The European Parliament and the Council of the European Union. Regulation - 2017/745 - EN - medical device regulation. *EUR-Lex* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745 (2017).
37. Center for Drug Evaluation & Research. Digital health technologies for remote data acquisition in clinical investigations. *U.S. Food and Drug Administration* https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digit

al-health-technologies-remote-data-acquisition-clinical-investigations (2023).

38. International Organization for Standardization. IEC 62366-1:2015. *ISO* https://www.iso.org/standard/63179.html (2021).
39. International Organization for Standardization. IEC/TR 62366-2:2016. *ISO* https://www.iso.org/standard/69126.html (2024).
40. International Organization for Standardization. ISO 9241-210:2019. *ISO* https://www.iso.org/standard/77520.html (2019).
41. International Organization for Standardization. ISO 20282-1:2006. *ISO* https://www.iso.org/standard/34122.html (2020).
42. International Organization for Standardization. ISO/TS 20282-2:2013. *ISO* https://www.iso.org/standard/62733.html (2024).
43. Digital Medicine Society. Library of Human Factors Resources for Digital Health Technologies. *DATAcc by DiMe* https://datacc.dimesociety.org/resources/interactive-index-of-human-factors-resources-for-digital-health-technologies-dhts-2/ (2023).
44. International Organization for Standardization. ISO 14971:2019. *ISO* https://www.iso.org/standard/72704.html (2019).
45. Zhou, Y. *et al.* Missing data matter: an empirical evaluation of the impacts of missing EHR data in comparative effectiveness research. *J. Am. Med. Inform. Assoc.* **30**, (2023).
46. Henni, S. H., Maurud, S., Fuglerud, K. S. & Moen, A. The experiences, needs and barriers of people with impairments related to usability and accessibility of digital health solutions, levels of involvement in the design process and strategies for participatory and universal design: a scoping review. *BMC Public Health* **22**, (2022).
47. Duffy, A., Christie, G. J. & Moreno, S. The challenges toward real-world implementation of digital health design approaches: narrative review. *JMIR human factors* **9**, (2022).
48. Office of Regulatory Affairs. Design controls. *U.S. Food and Drug Administration* https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/inspection-guides/design-controls (2023).
49. Schaefer, S. E., Van Loan, M. & German, J. B. A feasibility study of wearable activity monitors for pre-adolescent school-age children. *Prev. Chronic Dis.* **11**, (2014).
50. Davis, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly.* **13,** 319-340 (1989).
51. Shachak, A., Kuziemsky, C. & Petersen, C. Beyond TAM and UTAUT: future directions for HIT implementation research. *J. Biomed. Inform.* **100**, (2019).
52. Virzi, R. A. Refining the test phase of usability evaluation: how many subjects Is enough? *Hum. Factors* **34,** 4; 10.1177/001872089203400407 (1992).
53. Lewis, J. R. Sample sizes for usability studies: additional considerations. *Hum. Factors* **36**, (1994).

54. Digital Medicine Society (DiMe). DATAcc by DiMe resources. *DATAcc by DiMe* https://datacc.dimesociety.org/resources/ (accessed Feb 21, 2024).

55. Shoemaker, S. J., Wolf, M. S. & Brach, C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. *Patient Educ. Couns.* **96**, (2014).

56. FDA-NIH Biomarker Working Group. Glossary. in *BEST (Biomarkers, EndpointS, and other Tools) Resource* https://www.ncbi.nlm.nih.gov/books/NBK338448/#IX-C (U.S. Food and Drug Administration and National Institutes of Health, 2021).

57. Nielsen, J. Usability 101: introduction to usability. *Nielsen Norman Group* https://www.nngroup.com/articles/usability-101-introduction-to-usability/ (2012).

58. Center for Devices & Radiological Health. Applying human factors and usability engineering to medical devices. *U.S. Food and Drug Administration* https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices (2019).

59. U.S. Food and Drug Administration. Glossary of terms on clinical trials for patient engagement advisory committee meeting. *U.S. Food and Drug Administration* https://www.fda.gov/media/108378/download (2017).

60. Center for Devices & Radiological Health. General/specific intended use - guidance for industry. *U.S. Food and Drug Administration* https://www.fda.gov/regulatory-information/search-fda-guidance-documents/generalspecific-intended-use-guidance-industry (2020).

61. International Organization for Standardization. ISO 13485:2016. *ISO* https://www.iso.org/standard/59752.html (2020).

62. U.S. Food and Drug Administration. Human factors studies and related clinical study considerations in combination product design and development- draft guidance for industry and FDA staff. *U.S. Food and Drug Administration* https://www.fda.gov/files/about%20fda/published/Human-Factors-Studies-and-Related-Clinical-Study-Considerations-in-Combination-Product-Design-and-Development.pdf (2016).

63. Schnall, R., Higgins, T., Brown, W., Carballo-Dieguez, A. & Bakken, S. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mHealth technology use. *Stud. Health Technol. Inform.* **216**, 467 (2015).

64. Pillalamarri, S. S., Huyett, L. M. & Abdel-Malek, A. Novel bluetooth-enabled tubeless insulin pump: a user experience design approach for a connected digital diabetes management platform. *J. Diabetes Sci. Technol.* **12**, (2018).

65. Preusse, K. C., Mitzner, T. L., Fausset, C. B. & Rogers, W. A. Older adults' acceptance of activity trackers. *J. Appl. Gerontol.* **36**, (2017).

66. Nielsen, J. Usability inspection methods. *Nielsen Norman Group* https://www.nngroup.com/books/usability-inspection-methods/ (1994).

67. Standoli, C. E. *et al.* A smart wearable sensor system for counter-fighting overweight in teenagers. *Sensors* **16**, (2016).

68. Stubberud, A., Tronvik, E., Olsen, A., Gravdahl, G. & Linde, M. Biofeedback treatment app for pediatric migraine: development and usability study. *Headache* **60**, (2020).

69. Domingos, C., Costa, P., Santos, N. C. & Pêgo, J. M. Usability, acceptability, and satisfaction of a wearable activity tracker in older adults: observational study in a real-life context in northern Portugal. *J. Med. Internet Res.* **24**, (2022).