Validating Novel Digital Clinical Measures

DATAcc by DIGITAL MEDICINE SOCIETY

Digital Health Measurement Collaborative Community

Build Your Study: Novel Digital Clinical Measure Validation Planning

How to use this guide

Analytical validation is a crucial step in evaluating any digital clinical measure and is a key component of the V3+ framework. This guide will help you design the analytical validation study for your novel digital clinical measure, using the principles explained in the Interactive Guide to Validating Novel Digital Clinical Measures.

This document guides you step-by-step through six short sections, prompting you to provide appropriate information to optimize your selection of:

- ✔ Reference measures
- ✔ Choice of statistical analysis methodology
- ✔ Strategy for interacting with regulators

You'll find additional context and resources on the project landing page to support this process.

A novel digital clinical measure captures an outcome that has never previously been assessed, either by a digital or non-digital tool.

Table of contents

Section 1. Define objectives and evaluate existing evidence

STOP!

Do you have evidence to support that the novel digital clinical measure you are developing is meaningful?

If so, please proceed to planning your analytical validation study below. If not, please use DiMe's Digital Measures That Matter framework to guide you in selecting and developing a measure that will be meaningful for patients before proceeding.

Throughout this section, the resources from the Measurement Toolkit in DiMe's *Playbook* will help you answer the questions and provide the necessary evidence to ensure you are ready to create your analytical validation study.

Define your digital clinical measure of interest and the sensor-based digital health technology (sDHT) used to capture it.

What is the meaningful aspect of health (MAH) you intend to capture?

Example: "Our MAH is good quality of sleep in people with atopic dermatitis."

Which concept of interest (COI) do you intend to capture?

Example: "Our COI is nocturnal scratch."

What is the digital clinical measure (outcome) you intend to capture? List all properties that are important, such as its units, measurement interval, amplitude, count, start/end triggers, minimum duration, etc.

Example: "Our outcome is total scratching time (in seconds), defined as an action of rhythmic and repetitive skin contact movement performed during a delimited period of intended and actual sleep within the total sleep opportunity."

Which sDHT do you plan to use to capture the outcome, and what sample-level data does it output?

Example: "The sDHT used to capture it is [manufacturer / model of sDHT], a wrist-worn accelerometer sampling at 50 Hz."

Define and document your use context, including your population of interest and your intended use environment.

Examples: "Our context of use is an adult patient population aged 18 to 64 with eczema and reporting sleep disturbance. Our intended use environment is in a patient's regular sleeping arrangements in their home.

"Our context of use is to monitor daily calorie expenditure to compare with calorie consumption in a weight loss study for adults aged 18 to 45."

Document the algorithm and its requirements.

The algorithm requirements outline the characteristics of the input data, including minimum sampling frequencies, duration, sensor performance requirements, and any processing steps or calculations needed between the input data and the processed signal, such as analog to digital conversion of the input data, or details of any classification model used including all intermediate steps and how the model was trained. Also, draw on elements of the technical specification when answering this question.

Provide evidence for verification of the sDHT's sensor.

Verification is the evaluation of sensor accuracy, precision, consistency, and uniformity.

This may include manufacturer documentation or peer-reviewed manuscripts and does not require you to conduct your own verification study if you are not the technology manufacturer. Evidence provided should be easily understandable to non-technologists, and should include performance specifications for the integrated hardware, output data specifications, and software system tests.

Verification should be completed according to the principles outlined in the V3+ Framework. Please also use this table from Goldsack et al. (2020) to guide you in documenting, compiling and presenting the necessary evidence.

Describe your approach for assessing sufficient evidence of analytical validation for your measure by setting clear criteria.

Setting clear criteria to define when analytical validation is considered "achieved" for your digital clinical measure is an important step in your study planning.

Your approach for establishing sufficiency of analytical validation will depend on many factors, including existing evidence about your measure of interest or similar measures, and your intended context of use.

For measures that have at least some established evidence, conducting a literature review is a good place to start. If a good reference measure exists (see Section 2 for more details on how to define "good"), a numerical accuracy threshold can be set as

a pass/fail criteria. Consider both the expected accuracy of the reference measure and what would be considered clinically meaningful accuracy for your digital clinical measure of interest.

For novel measures with very little or no established evidence, defining a pass/fail threshold may be difficult or impossible. As above, it is important to consider the level of accuracy necessary for your digital clinical measure of interest to measure something clinically meaningful. Measure aspects like repeatability, reliability, and the ability to detect change over time should also be considered. Adapting accuracy thresholds identified in literature or by key opinion leaders may also be appropriate.

The environment in which analytical validation is conducted may also affect selection of your evidentiary criteria. Section 4 will guide you through these important considerations.

Note that conducting a pilot study is not a requirement to complete this step. For more background, refer to the "How much validation is "enough"?" section in Goldsack et al. (2020).

First, summarize the known evidence that may help you decide on an appropriate reference measure and test criteria, both within and outside your context of use. Explain how the identified evidence supports your analytical validation plan. Please use this table from Goldsack et al. (2020) to guide you in documenting, compiling, and presenting the known evidence.

Example: "We require that our digital clinical measure exhibits strong agreement (exceeding a [your pass/fail criteria]) with at least one high-quality reference measure when analyzed in the

[intended use environment] and [population of interest]. We adapted this approach from [reference to literature] that validated a similar measure."

Second, consider the available evidence and your best assessment of how the evidence meets your pass/fail criteria, if possible. Describe where evidence may be lacking.

Examples of evidence gaps may include:

"Analytical validation that meets the proposed pass/fail criteria exists for this measure in a lab environment, but not in the intended use environment."

"Analytical validation exists in our intended use environment but does not meet our proposed pass/fail criteria."

"No prior analytical validation evidence exists for this measure."

"Strong analytical validation evidence exists for this measure in a population of women over 40, but no analytical validation exists for our population of interest of women aged 25-34, and we believe this may affect the accuracy of our measure for [list reasons].

STOP!

Before proceeding, consider meeting with regulators for feedback on the existing evidence, its limitations, and the proposed analytical validation study design.

Early engagement with regulators is critical when regulatory acceptance and qualification are required.

To help choose your best pathway for interacting with the US Food and Drug Administration when developing your novel digital clinical measure, please use the Regulatory Quick Start Guide in DiME's *Playbook* and resources from DiME's Digital Health Regulatory Pathways communication toolkit, such as Engagement Pathways to Communicate with U.S. Regulators (FDA - Food and Drug Administration).

✔ **Check**

Does the identified evidence robustly establish that the combination of your digital clinical measure of interest and your sDHT requires no further analytical validation?

That is to say, the identified evidence robustly establishes that the algorithm correctly captures your physiological measure or behavioral construct of interest in your intended context of use.

☐ Yes → Skip to Section 6.

☐ No → Continue to Section 2.

Section 2. Identify appropriate reference measures for the analytical validation of your measure of interest

This section, and those following, will help you design the most appropriate analytical validation study for your measure of interest if you identified that need in Section 1.

Not all reference measures are created equally. Choose reference measures using the following hierarchy, with the rigor of the reference measure category decreasing from left-to-right:

Defining → **Principal** → **Manual** → **Reported**

If you are unfamiliar with the types of reference measures and their hierarchy, please refer to Stage 2 of the Interactive Guide to Validating Novel Digital Clinical Measures.

Note that throughout, we do not necessarily refer to a single reference measure. Although it may be appropriate to use a single reference measure, it will depend on the existing evidence and other information you collated as part of Section 1. Specifically, the need for multiple reference measures may increase as reference measures of lower rigor are selected. It may also depend on your context of use and how that affects the performance of the algorithm.

Evaluate available defining reference measures.

A defining reference is an objective measure that emerges when a physiologic process or behavioral construct depends on the technology used to capture it to such an extent that it sets the medical definition for that process or construct. Defining measures always have a standards document issued by a respected professional body. The units of a defining reference are directly comparable to the digital clinical measure of interest.

Directly comparable means the units are either the same, or can be translated for the purposes of comparison (e.g., via calibration).

Start by listing all defining measures that **may** be appropriate to analytically validate your digital clinical measure of interest.

Although defining measures are of the highest rigor in the hierarchy, the availability of one or more defining measures for your measure of interest does not necessarily mean they are an appropriate option.

Below, note the defining measures you believe to be appropriate for your study, and justify why. Pay particular attention to factors like:

- Complexity of measurement
- Applicability for your intended environment of use
- Potential safety considerations

✔ **Check**

One or more appropriate defining reference measures exist, and I will not need additional reference measures to meet my criteria for analytical validation.

To decide if the reference measures identified so far are sufficient, refer to your answers in Section 1; specifically, any evidence gaps between your analytical validation criteria and your compiled evidence. If an analytical validation study for your digital clinical measure of interest against your chosen reference measure(s) is likely to meet your criteria for sufficient evidence, then you can answer Yes to this question. Otherwise, or when in doubt, answer No.

☐ Yes → Skip to the end of Section 2.

☐ No → Continue below.

Evaluate available principal reference measures.

A principal reference measure is a direct and objective measure of the physiologic process or behavioral construct of interest. Principal reference measures may have a standard document issued by a respected professional body. The units of a principal reference are directly comparable to the digital clinical measure of interest.

List all principal measures that **may** be appropriate to analytically validate your digital clinical measure of interest.

Principal measures rely on objective measurement and are the best option if defining measures are not available or practical. However, the existence of a principal measure does not necessarily mean it is suitable for your study.

Below, note the principal measures you believe to be appropriate for your study, and justify why. Pay particular attention to factors like:

- Complexity of measurement
- Applicability for your intended environment of use
- Potential safety considerations

✔ **Check** **One or more appropriate principal reference measures exist, and I will not need additional reference measures to meet my criteria for analytical validation.**

To decide if the reference measures identified so far are sufficient, refer to your answers in Section 1; specifically, any evidence gaps between your analytical validation criteria and your compiled evidence. If an analytical validation study for your digital clinical measure of interest against your chosen reference measure(s) is

likely to meet your criteria for sufficient evidence, then you can answer Yes to this question. Otherwise, or when in doubt, answer No.

☐ Yes → Skip to the end of Section 2.

☐ No → Continue below.

Evaluate available manual reference measures.

A manual reference measure relies on the measurement, observation, or perception of a physiological process or behavioral construct by a trained healthcare professional, with or without equipment or technology. The units of a manual reference are directly comparable to the digital clinical measure of interest.

Start by listing all manual measures that **may** be appropriate to analytically validate your digital clinical measure of interest.

Manual reference measures introduce subjectivity into the measurement process and are open to factors such as observer bias. Nevertheless, they may be the best option once defining and principal measures have been considered.

Below, note the manual measures you believe to be appropriate for your study, and justify why. Pay particular attention to factors like:

- Complexity of measurement
- Applicability for your intended environment of use
- Interrater variability and observer bias

✔ Check

One or more appropriate manual reference measures exist, and I will not need additional reference measures to meet my criteria for analytical validation.

To decide if the reference measures identified so far are sufficient, refer to your answers in Section 1; specifically, any evidence gaps between your analytical validation criteria and your compiled evidence. If an analytical validation study for your digital clinical measure of interest against your chosen reference measure(s) is likely to meet your criteria for sufficient evidence, then you can answer Yes to this question. Otherwise, or when in doubt, answer No.

☐ Yes → Skip to the end of Section 2.

☐ No → Continue below.

Evaluate available reported reference measures.

A patient-reported reference measure is based on a report that comes directly from a patient about the status of their health condition, while an observer-reported reference measure is based on a report from another individual based on observable signs, events, or behaviors related to a patient's health condition. The units of a manual reference are ideally directly comparable to the digital clinical measure of interest. When that is not possible, refer to the Simulation Toolkit for Digital Clinical Measure Validation for statistical considerations.

List all reported measures that **may** be appropriate to analytically validate your digital clinical measure of interest.

Reported reference measures are ranked below manual measures as they cannot mitigate measurement variability by creating an average or consensus measure. Carefully consider if a single reported reference measure will be sufficient to meet your evidence criteria established in Section 1.

Below, note the reported measures you believe to be appropriate for your study, and justify why. Pay particular attention to factors like:

- Complexity of measurement
- Applicability for your intended environment of use
- Interrater variability and observer bias

✔ **Check**

My chosen reference measures will allow me to conduct a study that will achieve my previously defined analytical validation criteria.

Refer to your answers in Section 1; specifically, any evidence gaps between your analytical validation criteria and your compiled evidence. If an analytical validation study for your digital clinical measure of interest against your chosen reference measure(s) is likely to meet your criteria for sufficient evidence, then you can answer Yes to this question. Otherwise, or when in doubt, answer No.

☐ Yes → Skip to Section 4.

☐ No → Continue to Section 3.

Section 3. Develop novel comparators or identify anchor variables to use as reference measures

This section is only required if you answered **"No"** to the question at the end of Section 2, meaning you have not been able to identify one or more reference measures that allow you to conduct analytical validation at a level of rigor appropriate for the context of use of your digital clinical measure.

If no suitable measures exist, consider developing a novel comparator to use as a reference measure. Novel comparators can be classified as *manual* (such as a manual annotation of a video to identify nocturnal scratch episodes) or *reported* (such as a self-report of fatigue through ecological momentary assessment).

Choose novel comparators using the following hierarchy, with the rigor of the reference measure category decreasing from left-to-right:

Manual	→	**Reported**

For more information on the hierarchy of novel comparators and additional considerations when developing novel comparators, please see Stage 2 of the Interactive Guide to Validating Novel Digital Clinical Measures.

Is developing a novel manual comparator appropriate and feasible for analytical validation of your digital clinical measure of interest?

A manual comparator relies on the measurement, observation, or perception of a physiological process or behavioral construct by a trained healthcare professional, with or without the use of equipment or technology. The units of a manual comparator are directly comparable to the digital clinical measure of interest.

Focus your response on items like a validation strategy for the novel comparator, and consider why this strategy would generate an appropriate reference measure for the context of use of your measure of interest.

☐ Yes → Complete the three boxes below. Then, proceed to the end of Section 3.

☐ No → Continue to the next question.

Describe the novel manual comparator(s) you are considering designing:

Describe how you intend to validate the novel comparator(s):

Justify why they are appropriate (or not) to assess analytical validity of your digital clinical measure of interest:

Is developing a novel reported comparator appropriate and feasible for analytical validation of your digital clinical measure of interest?

A patient-reported comparator relies on a report directly from a patient about the status of their health condition. An observer-reported reference measure is based on a report from another individual based on observable signs, events, or behaviors related to a patient's health condition. The units of a reported comparator are ideally directly comparable to the digital clinical measure of interest. When that is not possible, refer to the Simulation Toolkit for Digital Clinical Measure Validation for statistical considerations.

Focus your response on items like a validation strategy for your novel digital clinical measure, and consider why this strategy would generate an appropriate reference measure for the context of use of your measure of interest.

☐ Yes → Complete the three boxes below. Then, proceed to the end of Section 3.

☐ No → Continue below to identify anchor measures.

Describe the novel reported comparator(s) you are considering designing:

Describe how you intend to validate the novel comparator(s):

Justify why they are appropriate (or not) to assess analytical validity of your digital clinical measure of interest:

Identify anchor measures

When developing a novel comparator is impossible, anchor measures can provide a last-resort solution. However, this option should be avoided as it is not proper analytical validation and should only be considered if all other options were exhausted.

An anchor is any interpretable measure of a physiologic process or behavioral construct similar to the digital clinical measure of interest. Anchor measures generate data in units that are not directly comparable to those of your digital clinical measure, and therefore, analysis is limited to examining associations and correlations.

For more information on anchor variables and how to best incorporate them into your study, please refer to Stage 2 of the Interactive Guide to Validating Novel Digital Clinical Measures and FDA's Draft Guidance on Patient-Focused Drug Development.

Select suitable anchor measures that support analytical validation of your measure of interest.

Consider attributes shared with defining, principal, manual, or reported reference measures, and develop a strategy for evaluating correlations and associations, considering your use context. It is highly recommended to use multiple anchor measures to strengthen any assertions that the algorithm measures what it is intended to measure.

Examples: MDS-UPDRS Motor Assessment Item 3.15 was selected as an anchor measure for our sDHT-derived measure of postural tremor (in units of acceleration). To avoid perpetuating the known limitations of this methodology by seeking to align our measure as closely as possible with the clinician's visual assessment of tremor amplitude, we selected further anchor measures, namely the PRO-PD rating scale.

Items from the PHQ-9 and NOSE assessment related to tiredness, energy, and sleep were chosen as anchor measures for our digital clinical measure of daily function, assessed through sleep variables.

✔

Reminder

If you have reached this stage, you **must** have chosen appropriate reference measures, novel comparators, or anchor measures that will allow you to conduct a study to meet your analytical validation criteria defined in Section 1.

Only proceed (to Section 4) if this is true.

Section 4. Assess the impact of your intended use environment on the measure performance

Ideally, your analytical validation study is conducted in the intended use environment but this may not always be feasible or sufficient, in which case the study can be conducted in the lab. In some cases, analytical validation may need to be split across in-lab and the intended use environment to meet your study objectives.

In the latter case, not all components must necessarily be conducted in both environments to meet your analytical validation criteria. It may be possible to reach sufficient evidence of validation with an extensive in-lab study for one aspect of your measure, while another aspect may require validation in your intended use environment, or validation in both environments. Please refer to Mahadevan et al. (2021) for an example of a successful approach of this type.

Refer to Stage 3 of the Interactive Guide to Validating Novel Digital Clinical Measures for a helpful flowchart to guide your decision making on this topic.

Assess the impact of your intended use environment.

First, briefly assess if you can conduct analytical validation in your intended use environment using the most rigorous identified reference methodology:

✔ **Check**

Is it possible to conduct the analytical validation study in the intended use environment using the most rigorous identified reference methodology?

☐ Yes → Your study can be conducted entirely in the intended use environment. Continue to Section 5 to design your study plan.

☐ No → Continue below.

✔ **Check**

Is your digital clinical measure expected to differ substantially, in accuracy or precision, when the sDHT is implemented in a lab setting as opposed to in your intended use environment?

☐ Yes → Continue below.

☐ No → Your study can be conducted entirely in the lab environment. Continue to Section 5 to design your study plan.

Is the expected performance difference likely due to environmental conditions impacting the sDHT sensor?

Example: An investigator is considering using a wrist-worn sDHT to capture heart rate in extremely high heat and is concerned that the signal quality may deteriorate in these conditions.

☐ Yes → Check verification evidence of the sDHT sensor in Section 1. If this evidence does not support sensor verification in the context of environmental conditions for your use, obtain further verification and add this to your answer in Section 1. Then continue to the next question.

☐ No → Continue to the next question.

Is the expected performance difference likely due to usability considerations?

Example: An investigator is considering using an EEG headband to capture sleep efficiency and is concerned that patients may place the headband incorrectly when using the sDHT at home unsupervised, leading to excessive signal artifacts.

☐ Yes → Usability validation should be completed according to the principles outlined in the V3+ Framework. Resources to help you design a strong usability validation plan aligned with the V3+ principles can be found on DiMe's V3+ resource hub. Then continue to the next question.

☐ No → Continue to the next question.

Are there remaining concerns that algorithm performance is likely to differ substantially when used in-lab *versus* the intended use environment(s)?

Example: An investigator is considering using an adhesive patch sDHT to measure overnight SpO_2 in children but plans to recruit adults for an analytical validation study involving arterial blood draws under hypoxic conditions because the ethical/IRB-approval process is more straightforward in this population.

☐ Yes → Your study requires a combination of in-lab and intended use environment validation activities. Proceed with in-lab analytical validation using the highest-order reference measure, novel comparator, or anchor measures, and capture supplementary analytical validation in the intended use environment(s) using a lower-order reference measure, novel comparator, or anchor measure(s). Continue to Section 5 to design your study plan.

☐ No → Your study can be conducted entirely in the intended use environment. Continue to Section 5 to design your study plan.

✔ **Check**

Are your chosen reference measures appropriate for conducting analytical validation of your measure in your intended use environment?

Go back to Section 2 or Section 3 and look at the factors you defined for your reference measures:

- Are the reference measures too complex or impractical to measure in the intended use environment?
- Are there potential safety considerations that preclude measuring a reference outside a lab environment?
- If a reference measure requires an observer, will the measurement environment potentially affect the accuracy of that observation?

☐ Yes → Briefly justify why your chosen reference measures are **appropriate** in your intended use environment.

☐ No → Briefly justify why your chosen reference measures are **inappropriate** in your intended use environment.

Then, return to Section 2 and use the reference measure hierarchy to select additional reference measures for your intended use environment, justifying their appropriateness. If suitable reference measures cannot be identified, proceed to Section 3 to develop novel comparators or identify anchors.

Section 5. Design a study plan

In this section, you will design a study plan for analytical validation of your digital clinical measure of interest in your intended use environment, in the lab, or in a combination of both environments, depending on your answers in Section 4.

When planning your study, consider whether the units of your digital clinical measure of interest are directly comparable to those of your reference measure. This impacts the selection of statistical methods, your choice of agreement statistics, and whether additional techniques for assessing the validity of a measure should be used.

Directly comparable means the units are either the same, or can be translated for the purposes of comparison (e.g., via calibration).

Statistical methods and agreement statistics that may be suitable for your study if the units are directly comparable include, but are not limited to: sensitivity and specificity, positive and negative predictive value, receiver operating characteristic curves, Kendall's τ rank distance (with or without item-weighting), intraclass correlation coefficients for absolute agreement between two raters, and the concordance correlation coefficient. Also consider whether the data you are collecting is categorical/ordinal, or continuous, when choosing your methods and agreement statistics.

Statistical methods and agreement statistics that may be suitable for your study if the units are **not** directly comparable include, but are not limited to: The Pearson correlation coefficient, or simple/multiple linear regression models (using R^2 or adjusted R^2 as the agreement statistic). When pursuing convergent validity, a confirmatory factor analysis model, using the factor correlation as the agreement statistic, may be appropriate. When pursuing known-groups validity, calculating effect sizes using Cohen's d may be an appropriate agreement statistic.

For more information on these topics to aid you in answering the following questions, please see the Simulation Toolkit for Digital Clinical Measure Validation.

In each question, your answers should consider the intended use environment, the lab setting, or a combination of both environments, based on your answers in Section 4 and your choice of reference measures, comparators or anchors in Section 2 and Section 3.

Develop a fit-for-purpose statistical strategy

First, describe your strategy to maximize **data completeness** in both your digital clinical measure and your reference measure(s). This may include technical considerations, or social considerations such as a patient engagement strategy where an investigator reminds participants the day before the beginning of the sDHT wear

period via phone call. Carefully consider how your study design may lead to patterns of data missingness in the digital clinical measure and the reference measure, and what effect any missingness patterns may be expected to have on your analysis results:

Second, describe your strategy for maximizing **temporal coherence** between your digital clinical measure and your reference measure. Temporal coherence describes the similarity between the time periods of data collection for two measures. Poor temporal coherence between measures may decrease the values estimated with agreement statistics, because a participant's meaningful aspect of health assessed by the measures may have changed or fluctuated over time.

When using a reference measure with **non-directly comparable units** (such as a reported reference measure, comparator, or anchor) and a daily recall period, assessing the reference measure on the same days as the digital measure is recommended. In the case of a multi-day recall period, assessing the reference measure at the end of the period of digital measure data collection, and collecting digital measure data on each day of the recall period, is expected to increase temporal coherence.

Third, describe your strategy for maximizing **construct coherence** between your digital clinical measure and your reference measure. Construct coherence is the level to which your digital clinical measure and your reference measures assess the same underlying concept or latent construct. Poor construct coherence is likely to lead to weak or non-meaningful relationships between measures, no matter the statistical methods employed.

Fourth, describe the **statistical methods** to be used for assessing agreement between your digital clinical measure and each reference measure, comparator, or anchor, including all estimates or statistics to be calculated:

Provide an overview of additional study plan details needed to perform analytical validation in the required environments.

This includes, but is not limited to, factors such as population size, participant recruitment and enrolment strategy, participant inclusion/exclusion criteria, study sites or clinical centers to be used, the duration of a participant's involvement in your study, and more details on your study procedures themselves.

✔ **Check**

Is the strategy you have developed above appropriate for conducting analytical validation of your measure in the environments required by your answers in Section 4?

☐ Yes → Briefly justify why your strategy is appropriate for each required environment. Then continue below.

☐ No → Update your strategy above, to ensure that it is appropriate for each required environment. Then continue below.

STOP!

Before proceeding, consider meeting with regulators to present your complete study plan (including your choice of reference measures).

Continued engagement with regulators is critical when regulatory acceptance and qualification are required.

To help choose your best pathway for interacting with the US Food and Drug Administration when developing your novel digital clinical measure, please use the Regulatory Quick Start Guide in DiMe's *Playbook* and resources from DiMe's Digital Health Regulatory Pathways toolkit, such as Engagement Pathways to Communicate with U.S. Regulators (FDA - Food and Drug Administration).

Continue to Section 6.

Section 6. Create an analytical validation evidence dossier for further regulatory interaction and submission, combining all supporting evidence and information

At this stage, you can finalize your complete study plan, conduct your study, and perform all data analyses.

Your study plan should be included in a dossier that compiles all your evidence and documentation supporting the analytical validation of your digital clinical measure. As your study progresses, use the checklist below to help manage the development of your evidence dossier.

Preliminary data and study objectives

- ☐ Your digital clinical measure of interest and the sDHT being used to capture it.
- ☐ Context of use, including your intended population of interest and intended use environment.
- ☐ The algorithm and its requirements.
- ☐ Evidence for the verification of the sDHT's sensor.
- ☐ The pass/fail performance criteria for what constitutes sufficient evidence of analytical validation for your measure.
- ☐ The existing evidence of analytical validation of your digital clinical measure in your context of use.

Choice of reference measures

- ☐ **Identify** your selected reference measure(s) as defining, principal, manual, reported, novel, or anchored.
- ☐ **Justify** the choice of your reference measures, including their appropriateness for in-lab validation, intended use environment validation, or both.
- ☐ **Explain** the rationale for passing over a higher-ranked reference measure in the hierarchy favoring a less rigorous reference, comparator, or anchor.
- ☐ **Provide** evidence of the development and validation of any novel comparators used.

Analytical validation activities and results

- ☐ Study plan for conducting analytical validation of your measure of interest (in the lab, in the intended use environment, or both, based on your choices in Section 4 and Section 5).
- ☐ [If analytical validation in your intended use environment was not conducted] Rationale for why analytical validation in the intended use environment was not required.
- ☐ Analysis of agreement between the digital clinical measure data and reference measure data using your chosen statistical strategy.

Additional materials

- ☐ Code used in the analysis
- ☐ Relevant data sets
- ☐ Other pertinent materials not covered above