IntuitionLabs

INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

FDA Digitally Derived Measures in Clinical Trials

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-09-19 · fda digitally derived measures · digital endpoints · clinical trials · digital biomarkers · sensor-based endpoints · analytical validation · clinical validation · digital health technologies · change control

Original article: https://intuitionlabs.ai/articles/fda-digital-measures-evidence-matrix

In this article

1. Executive Summary
2. Introduction and Background
3. Definitions and Endpoint Taxonomy
4. Key Changes
5. Evidence Matrix by Endpoint Role
6. Implementation Considerations and Process Changes
7. Data Analysis and Evidence
8. Worked One-Week Stride-Length DDM (Hypothetical Example)
9. Implications and Future Directions
10. Frequently Asked Questions (FAQs)
11. Conclusion

Executive Summary

The U.S. Food and Drug Administration's **August 2026** cross-center paper changes the practical question for digitally derived measures (DDMs). The decision is not whether a wearable, sensor, or algorithm is novel. It is whether a defined measure is fit for a defined role in a defined trial. FDA describes a DDM as a measure derived from data collected using a digital health technology (DHT), and says it may be a **clinical outcome assessment (COA), biomarker, or component of a multicomponent endpoint** (fda.gov). The endpoint classification determines what the output means, while the endpoint hierarchy determines how much uncertainty the development program can tolerate.

The required traceability chain runs from **meaningful aspect of health**, through target population, concept of interest, context of use, DHT verification, DDM analytical validation, clinical validation, usability, and operational error controls. Verification asks whether the sensing system meets specifications. Analytical validation asks whether the algorithm produces an accurate measure against an appropriate reference. Clinical validation asks whether the measure captures the intended concept in its context of use (nature.com). FDA's pivotal-primary example expects **prospective validation and prespecified performance thresholds** (fda.gov). By contrast, CTTI recommends early introduction as an exploratory endpoint, where evidence can be accumulated without making the DDM carry the confirmatory claim (ctti-clinicaltrials.org).

The evidence base also shows why operational controls are part of validity. A review of **75 randomized trials and 10,822 patients** found only **20 trials, or 26%**, reported validity, reliability, and responsiveness information, while **51 trials, or 68%**, did not mention incomplete device-derived outcome data (link.springer.com) (link.springer.com) (link.springer.com). A separate review found **25 distinct adherence definitions** among 37 technologies that categorized adherence (jmir.org). These are design problems, not merely post hoc data-cleaning problems.

For sponsors, the actionable approach is an endpoint-role evidence matrix, prospective context-specific acceptance criteria, explicit missing-data estimands, locked and traceable software versions, and a bridging rule for every material change. There is no defensible universal accuracy or missingness threshold. DiMe

explicitly states there is **no universal pass threshold** for analytical validation (navigator.dimesociety.org), and the 2025 implementation paper says tolerated missingness depends on the DDM's purpose (nature.com).

Introduction and Background

Digitally derived measures can capture function, physiology, or experience repeatedly outside periodic site visits. FDA's **2026 DDM paper** was jointly published by its medical product centers as a white paper (fda.gov). It synthesizes existing guidance rather than creating a single new qualification pathway. Its practical contribution is to connect patient relevance, measurement science, software assurance, and trial operations into one fit-for-purpose argument.

That integration matters because a DHT is not itself an endpoint. The endpoint is the precisely defined trial variable. The NIH-FDA BEST resource defines an endpoint as **a precisely defined variable intended to reflect an outcome of interest** (ncbi.nlm.nih.gov). A wrist device, for example, can yield raw acceleration, activity counts, walking bouts, sleep estimates, or a composite score. Each output needs its own concept, algorithm specification, aggregation rule, and evidence.

The distinction is especially important for sponsor clinical scientists, digital biomarker leads, biostatisticians, data-management teams, contract research organizations, and DHT vendors. Their shared decision is whether a proposed sensor-based endpoint is mature enough for **exploratory, secondary, or primary and label-supporting use**. IntuitionLabs is adjacent to this decision as a life-sciences implementation and advisory consultancy, not as a DHT endpoint vendor. Its public service description emphasizes data pipelines, integration, warehousing, and business intelligence (intuitionlabs.ai). That perspective is relevant to traceability and operationalization, but it is not evidence that a particular DDM is valid.

This report turns the FDA paper into an implementation artifact. It first separates the key terms, then organizes evidence by endpoint role, operationalizes the evidence ladder, and applies the method to a hypothetical one-week stride-length DDM.

Definitions and Endpoint Taxonomy

DHT, DDM, endpoint, COA, and biomarker

Digital health technology describes the technical system. **Digitally derived measure** describes the output derived from data the system collects. **Endpoint** describes how a precisely specified variable is analyzed to answer a trial objective. These layers should remain separate in the protocol, statistical analysis plan, data-management plan, and vendor specification.

A DDM can occupy different regulatory categories:

- **Clinical outcome assessment:** a measure of how a person feels, functions, or survives, reported by a clinician, patient, observer, or performance task. FDA lists those reporting modes (fda.gov), and BEST recognizes four corresponding COA types (ncbi.nlm.nih.gov).

- **Biomarker:** a measured characteristic indicating a biological process or response. FDA defines it as a measured characteristic used as an indicator (fda.gov), while BEST distinguishes it from a measure of how a person feels, functions, or survives (ncbi.nlm.nih.gov).
- **Multicomponent endpoint:** a within-participant combination of at least two components. FDA's DDM summary expressly includes measures used as part of multicomponent endpoints (fda.gov).
- **Supporting measurement:** a DDM may be informative without serving as a formal endpoint, for example as an adherence signal, stratification variable, or mechanistic readout. The claim must not outrun this designated role.

Calling every sensor output a digital biomarker creates two problems. First, it obscures whether the measure captures biological state or clinical function. Second, it encourages teams to validate a device in general instead of validating a specified output in a specified context. The correct noun is the role the measure actually plays.

The traceability chain begins with the patient

FDA defines a **meaningful aspect of health** as an aspect of feeling or functioning in daily life that matters to patients (fda.gov). The **concept of interest** is what the COA specifically measures (fda.gov). The **context of use** states the target population, trial setting, endpoint role, interpretation, and decision the score supports.

The sponsor should be able to trace the following chain without a semantic jump:

1. **Patient priority:** what feeling, function, or survival outcome matters?
2. **Target population:** in whom, at what disease stage, with which relevant capabilities and limitations?
3. **Concept of interest:** which specific behavior, function, physiology, or experience represents that priority?
4. **Measurement approach:** which observable signal and derived variable operationalize the concept?
5. **Context of use:** what trial phase, endpoint role, setting, duration, estimand, and decision will use it?
6. **Interpretation:** what direction and magnitude of change would be meaningful, and on what evidence?

The context should identify the target population early. FDA defines fit-for-purpose evidence as validation sufficient to support the context of use (fda.gov). Clinical validation is likewise context-specific. A technically accurate gait algorithm validated in healthy adults therefore does not automatically establish clinical validity in people with Parkinson disease, assistive-device use, or severe fatigue.

Key Changes

Endpoint hierarchy, not device novelty, sets the evidence burden

The most consequential shift is to organize the program around **decision consequence**. EMA's 2026 methodology guidance expresses the same risk logic: the greater a method's impact on regulatory decision-making, the more robust its evidence base should be (ema.europa.eu). Its context-of-use examples explicitly distinguish primary or secondary efficacy endpoints and safety monitoring.

The hierarchy is practical:

- **Exploratory use:** learn feasibility, distribution, adherence, sources of error, responsiveness, and candidate thresholds. CTTI recommends introducing DDMs as exploratory endpoints in early-phase trials (ctti-clinicaltrials.org).
- **Secondary use:** support a prespecified objective and multiplicity strategy. Evidence should be sufficiently mature to justify interpretation, missing-data handling, and directional claims.
- **Primary or label-supporting use:** carry a confirmatory conclusion, supported by prospective validation and prespecified performance thresholds.
- **Cross-program qualification:** support reuse within a defined scope. The EMA process can lead to either qualification advice or a CHMP qualification opinion (ema.europa.eu).

Verification and validation are separate claims

The evidence ladder contains distinct questions:

- **DHT verification:** does the sensor and acquisition system meet prespecified technical performance requirements under intended conditions?
- **DDM analytical validation:** does the algorithm transform the captured signal into the target metric accurately, precisely, and reliably against an appropriate reference?
- **Clinical validation:** does the DDM identify, measure, or predict the intended meaningful state or experience in the specified population and context?
- **Usability validation:** can intended users set up, wear, charge, synchronize, and troubleshoot the system well enough to obtain interpretable data? FDA expects evidence that users understand and can follow the instructions (fda.gov).

The V3 framework places sample-level sensor evaluation before algorithm evaluation (nature.com). It separately evaluates algorithms that convert sensor samples into physiological metrics (nature.com), then compares the algorithm-derived metric with an appropriate reference (nature.com). FDA also says a trial DHT should be verified and validated as fit for purpose (fda.gov). The V3+ extension adds usability and calls for quantitative pass or fail criteria to be specified in advance (nature.com).

Error analysis becomes part of endpoint design

Device settings, storage, transmission frequency, and environmental limits all affect missing-data risk. These operating conditions make failure modes part of the validity argument, not merely logistics.

Teams should distinguish at least five mechanisms:

- **Not worn:** participant choice, burden, skin discomfort, charging, or misunderstanding.
- **Worn but not measuring:** sensor placement, contact, battery, firmware, or environmental conditions.
- **Measured but not transferred:** phone pairing, local storage, network, synchronization, or server ingestion.
- **Transferred but not processed:** corrupt files, clock errors, unsupported version, or pipeline failure.

- **Processed but not analysis-ready:** insufficient wear window, implausible values, failed quality control, or missing covariates.

Each category has different implications for missing-at-random assumptions, remediation, and sensitivity analysis. Collapsing them into a single completeness percentage destroys useful causal information.

Evidence Matrix by Endpoint Role

Table 1 translates endpoint hierarchy into a minimum evidence posture. It is a planning matrix, not a universal regulatory checklist. “Minimum” means the evidence needed to justify that role in the stated context, not the lowest effort a team can document.

Evidence domain	Exploratory endpoint	Secondary endpoint	Primary or label-supporting endpoint
Meaningfulness and concept	Document patient or clinician rationale and a testable concept hypothesis.	Confirm content relevance in the target population and prespecify interpretation.	Demonstrate a complete chain from meaningful aspect of health to the proposed claim, with patient input and regulator alignment.
DHT verification	Bench evidence for core signal capture and known operating limits.	Verification across intended devices, placements, environments, and sites.	Locked specifications, production-equivalent units, lot or device variability, stress conditions, and prespecified pass criteria.
Analytical validation	Feasibility comparison with a defensible reference and error characterization.	Representative target-population study, subgroup analysis, reproducibility, and prespecified acceptance criteria.	Prospective, adequately powered validation independent of model development, covering the full intended range and critical subgroups. FDA says AI test data should be independent of training and tuning data (fda.gov).
Clinical validation	Association with relevant clinical measures and initial responsiveness estimates.	Evidence that the DDM measures the intended concept and supports the secondary interpretation.	Prospective evidence for the exact population, setting, aggregation, time window, estimand, and claim.
Usability and adherence	Identify use errors and burden; refine instructions and support.	Validate intended users, languages, sites, and remote workflows.	Demonstrate robust use under trial-like conditions and quantify residual use-error risk.
Missing data and statistics	Describe completeness by participant, device, day, and failure mode.	Prespecify usable-day rules, intercurrent-event handling, estimand, and sensitivity analyses.	Justify thresholds and assumptions with external or blinded internal evidence, simulate power impact, and prospectively lock rules.
Change control	Record all hardware, firmware, software, and algorithm versions.	Freeze analysis-critical versions; define impact-assessment and bridging triggers.	Lock the validated stack for pivotal collection or complete prospective bridging before pooled interpretation, with evidence that changes passed required tests (pages.nist.gov).

Evidence domain	Exploratory endpoint	Secondary endpoint	Primary or label-supporting endpoint
Regulatory interaction	Seek feedback when the concept or pathway is novel.	Align before the measure becomes decision-critical.	Obtain explicit, timely feedback on endpoint definition, evidence package, thresholds, SAP, and change plan.

The matrix prevents a common inversion: selecting a device first and then searching for a claim. Outcome selection should precede technology selection. When accepted clinical-benefit endpoints already exist, a novel DDM may be more useful as a complementary assessment than as an immediate replacement.

No cell implies a universal numerical threshold. The original V3 paper says a single analytical-accuracy threshold would be counterproductive because reference quality and intended use differ (nature.com). DiMe likewise states that analytical validation has no universal pass threshold (navigator.dimesociety.org). A one-centimeter error may be negligible for one concept and unacceptable for another. The sponsor must connect each criterion to clinical interpretation, statistical operating characteristics, or both.

Implementation Considerations and Process Changes

Build a traceability artifact, not a narrative memo

The evidence package should behave like a requirements system. Every claim maps backward to a specified endpoint, measure, algorithm, sensor signal, user action, and patient-relevant concept. Every test maps forward to a decision.

Table 2 shows the recommended traceability structure and the owner who should maintain each link.

Traceability node	Required artifact	Primary owner	Key review question
Meaningful aspect of health	Patient-input summary and conceptual model	Clinical science, patient engagement	Does the proposed measure reflect something that matters in daily life?
Concept of interest	Concept definition, boundaries, and expected change	Clinical science, COA or biomarker lead	Is the concept function, biology, behavior, or experience?
Context of use	Population, setting, role, time frame, interpretation, decision	Clinical and regulatory leads	Is the claim limited to the evidence-generating context?
DHT signal	Sensor specification, placement, sampling, clock, storage, environment	Vendor and digital technology lead	Can the system capture the source signal within specification?
DDM algorithm	Versioned code, features, preprocessing, quality control, aggregation	Data science and vendor	Can the exact output be reproduced from immutable source data?
Validation evidence	Verification, reference comparison, clinical and usability reports	Cross-functional evidence lead	Do studies match the intended population and use conditions?
Trial endpoint	Protocol definition, estimand, SAP, missing-data and sensitivity plan	Biostatistics and clinical science	Does the analysis answer the stated objective without changing the construct?

Traceability node	Required artifact	Primary owner	Key review question
Claim or decision	Interpretation guide and regulator feedback record	Regulatory and clinical leadership	Is every conclusion within the validated scope?

The table also clarifies vendor boundaries. Outsourcing does not outsource accountability. ICH E6(R3) places ultimate responsibility for trial-related activities and reliable data with the sponsor (database.ich.org). A sponsor therefore needs access to evidence and source-level metadata, not only a vendor summary.

Select reference measures and thresholds prospectively

The analytical-validation plan should define:

- **Measurand:** the exact quantity, unit, sampling interval, aggregation, and valid range.
- **Reference:** why the comparison method is appropriate for the construct and operating range.
- **Synchronization:** how the DHT and reference clocks, events, and analysis windows align.
- **Error metrics:** bias, precision, limits of agreement, classification metrics, or agreement measures appropriate to the output.
- **Acceptance criteria:** context-specific thresholds fixed before unblinding or final analysis.
- **Range coverage:** speeds, severities, behaviors, environments, placements, and artifacts expected in use.
- **Subgroups:** demographic and clinical strata that may influence signal quality or algorithm performance. Compare subgroup results against criteria specified in advance (navigator.dimesociety.org). FDA also recommends including participants across relevant demographic groups in performance studies (fda.gov).
- **Independence:** separation of algorithm development, tuning, and final validation data. Reusing one dataset can overestimate digital-endpoint utility (nature.com).

CLSI EP09 includes difference-plot and regression procedures for comparing quantitative measurement procedures and estimating bias (clsi.org). If no defensible reference exists, the claim and analysis may need to change. FDA recommends reporting agreement measures rather than accuracy when a reference standard is unavailable and cannot be constructed (fda.gov).

Quantify missingness at several denominators

A single “data completeness” number is insufficient. Calculate and visualize at least:

- **Participant level:** participants with enough evaluable data divided by randomized participants.
- **Device level:** devices producing valid data divided by devices deployed.
- **Day level:** valid participant-days divided by expected participant-days.
- **Window level:** valid analysis windows divided by scheduled windows.

- **Sample level:** accepted sensor samples divided by expected samples while the device should have been operating.
- **Transfer level:** source files ingested intact divided by files expected from the device.

For each participant, a transparent day-level rate is **valid days / expected days**, with numerator and denominator rules specified before analysis. Report its distribution, not just its mean. This discipline matters because one review found **25 unique adherence definitions** (jmir.org). Link every invalid day to a reason code and preserve the pre-quality-control count.

Control versions and bridge material changes

The minimum version record includes hardware model and revision, firmware, mobile application, operating system, configuration, algorithm, calibration, data dictionary, pipeline code, and analysis package. The 2025 implementation paper recommends the ability to freeze the algorithm before study start as a minimum technical requirement (nature.com). ISO 14155:2026 expressly covers clinical investigations involving Software as a Medical Device where relevant (committee.iso.org). W3C provenance provides a useful conceptual model by defining provenance as information about the entities, activities, and people involved in producing data (w3.org).

For every proposed update:

1. **Inventory the change:** identify affected components, outputs, metadata, and trial periods.
2. **Assess impact:** determine whether signal capture, preprocessing, quality control, output distribution, or interpretation could change.
3. **Classify materiality:** no impact, verification-only, analytical-validation impact, clinical-interpretation impact, or unknown.
4. **Choose evidence:** regression testing, bench verification, retrospective paired reprocessing, prospective bridging, or full revalidation.
5. **Decide deployment:** reject, defer, deploy prospectively with separation, or bridge and pool.
6. **Preserve auditability:** retain the old version, code, parameters, test evidence, deployment dates, and affected participant records.

Bridging may use paired reprocessing of retained data or new prospective evidence, depending on materiality. Subsequent computerized-system changes require risk-based validation. NIST's release model calls for evidence that every software change passed the required tests (pages.nist.gov).

Ask vendors for reproducible evidence

The sponsor's request package should include evidence generated under conditions that mirror intended use (navigator.dimesociety.org):

- **Specifications:** sensing modality, sampling, storage, battery, placement, operating limits, and device variability.
- **Algorithms:** exact version, preprocessing, features, model provenance, training data scope, and locked parameters.

- **Verification:** protocols, acceptance criteria, raw summaries, deviations, and production-unit comparability.
- **Analytical validation:** reference system, synchronization, population, range, subgroups, errors, and confidence intervals.
- **Clinical validation:** concept, population, setting, comparators, responsiveness, and interpretation.
- **Usability:** intended users, languages, tasks, observed use errors, instructions, training, and support model.
- **Operations:** data flow, timestamps, transfer retries, quality flags, incident handling, and disaster recovery.
- **Change control:** release history, notification period, impact assessment, bridging support, and end-of-support policy.
- **Auditability:** immutable source data, metadata, logs, access records, code traceability, and reproducible outputs.

DiMe procurement guidance calls for documentation covering verification, analytical validation, and clinical validation (dimesociety.org). This request should expose quantitative evidence, not only conclusions.

Data Analysis and Evidence

The published evidence shows both growth and uneven maturity. A ClinicalTrials.gov analysis estimated that connected digital product use grew at an approximately **34% compound annual rate from 2000 through 2017**, with more than **1,100 unique trials** using one in each of 2017 and 2018 (nature.com) (nature.com). DiMe's library reported **105 sponsors and 636 digital endpoints** as of April 2026 (dimesociety.org). These figures measure activity, not acceptance or adequacy.

Evidence quality is more mixed. A CTTI systematic review identified **275 feasibility publications**, with wearable sensors representing **67%** of technologies (nature.com) (nature.com). In the 75-trial review, **35 trials, or 47%**, used a biometric monitoring device for the primary outcome, yet **204 of 385 comparable outcome definitions, or 53%**, had not been prespecified in registries (link.springer.com) (link.springer.com).

Operational performance varies enough to make universal thresholds misleading. Across four neurological wearable studies, technical-error missingness averaged **10%**, with a **6% to 14% range** (pmc.ncbi.nlm.nih.gov). In one 408-participant study, daily adherence was **61% to 68%**, but only **53%** wore the sensor on both consecutive days (pmc.ncbi.nlm.nih.gov). Different populations, tasks, windows, and definitions limit direct comparison.

Reporting itself is a measurable risk. A usability scoping review found complete reporting of age, sex or gender, and race or ethnicity in only **14 of 83 studies, or 17%** (pubmed.ncbi.nlm.nih.gov). This result supports prospective denominator definitions, demographic performance plans, and standardized reporting rather than a single target completeness percentage.

Development datasets are often modest. A review of **262** DHT-enabled rare-disease trials found median enrollment of **48**, with an interquartile range of **25 to 94** (nature.com). DiMe reported that only **208 of 303** studies surviving its systematic-review process met its high-quality analytical-validation criteria (dimesociety.org). A 2005 to 2023 analysis found **99 trials, 0.075% of those assessed**, used digital biomarkers as endpoints (journals.sagepub.com). The combined signal is increasing experimentation alongside a persistent need for better prespecification, representativeness, and validation transparency.

Worked One-Week Stride-Length DDM (Hypothetical Example)

Consider a sponsor proposing **median free-living stride length over one week** as a secondary endpoint in a Parkinson disease trial. This is a hypothetical example, not an FDA-endorsed endpoint or an acceptable-threshold recommendation.

The meaningful aspect of health might be mobility in daily life. The concept of interest is not “accelerometer performance” or generic activity. It is sustained ambulatory function expressed through stride length during eligible walking bouts. NINDS guidance identifies feet or ankles as preferred locations for characterizing foot motion in Parkinson disease (commondataelements.ninds.nih.gov). It also advises clinical validation in the target Parkinson disease population rather than relying only on other populations.

Table 3 gives example acceptance-criterion fields. Every bracketed value is a protocol-development placeholder that the sponsor must justify from pilot data, clinical interpretation, power simulations, and regulator feedback.

Layer	Prospective specification	Example acceptance criterion	Evidence or action if unmet
DHT verification	Foot or ankle sensor, sampling rate, calibration, clock drift, environmental range	Bias within **[X unit]**, precision within **[Y unit]**, clock drift below **[Z]**	Correct hardware or configuration and repeat verification.
Analytical validation	Algorithm version, eligible bout definition, stride-length reference, synchronization	Mean bias and limits of agreement within **[context-specific bounds]** across the intended range	Refine the algorithm or narrow the context of use; validate on independent data.
Subgroup performance	Disease severity, assistive-device use, age, sex, body size, gait phenotype	Every critical subgroup meets **[prespecified bound]** or has a justified separate rule	Increase sample, stratify, exclude unsupported use, or revise interpretation.
Usability	Placement, charging, wear instructions, troubleshooting, remote support	Critical-task success at least **[X%]** and no unresolved critical use-error pattern	Redesign instructions, training, device, or support.
Day validity	Minimum wear duration, minimum eligible bouts, quality flags	A valid day requires **[X hours]** and **[Y bouts]**	Classify as missing with a reason code; do not silently impute a day.
Week validity	Number and distribution of valid days	A valid week requires **[X of 7 days]**, including **[weekday/weekend rule]**	Use prespecified estimand and sensitivity strategy.
Clinical validation	Relation to mobility construct, known-groups validity, responsiveness	Effect or correlation meets **[prospective criterion]** in target population	Retain as exploratory, revise concept, or collect additional evidence.

Layer	Prospective specification	Example acceptance criterion	Evidence or action if unmet
Version bridging	Paired processing of identical raw data plus prospective checks if needed	New and locked versions remain within **[equivalence bounds]**	Separate analysis periods, revalidate, or defer the update.

The reference system should match the measurand. Established gait-reference types include marker-based optical tracking, instrumented treadmills, and pressure-sensitive walkways (frontiersin.org). One treadmill validation reported a **1.7 cm** mean absolute stride-length difference, but its recordings were treadmill-based rather than overground (frontiersin.org) (frontiersin.org). That result is study-specific evidence, not a transferable threshold for this hypothetical endpoint.

A one-week window is plausible but still needs justification. One peer-reviewed mobility study collected wrist accelerometry over **7 days in 432 community-dwelling older adults** (academic.oup.com). This precedent supports feasibility only. It does not establish how many days, hours, or bouts are sufficient for Parkinson disease stride length.

Implications and Future Directions

The FDA paper favors a **modular evidence package**. Verification evidence can sometimes be reused for an unchanged sensor and configuration. Analytical validation may be leveraged when the algorithm, reference, population range, placement, and conditions remain sufficiently similar. Procurement should secure all three validation layers (dimesociety.org). Clinical validation is less portable because concept, population, setting, and endpoint interpretation define the context.

Expansion should therefore trigger a structured comparability review. V3+ notes that a new population may require repeating usability, analytical, and clinical validation (nature.com). MOBILISE-D's program illustrates a broader validation strategy, evaluating digital mobility outcomes for construct validity, predictive capacity, responsiveness, change detection, and minimal important difference (mobilise-d.eu). It also tested algorithms under real-life conditions across chronic obstructive pulmonary disease, multiple sclerosis, Parkinson disease, and proximal femoral fracture cohorts (mobilise-d.eu).

Regulatory interaction should occur before an endpoint becomes expensive to change. Available routes depend on product and purpose. FDA's ISTAND process offers nonbinding pre-submission consultations lasting **30 to 45 minutes** (fda.gov). EMA's pathway can lead to qualification advice or a CHMP qualification opinion (ema.europa.eu).

Operationally, future-ready programs will preserve raw data, standardized provenance, unambiguous timestamps, and enough metadata to reprocess measures transparently. IntuitionLabs describes its adjacent role as advisory guidance on digital transformation, AI adoption, and technology roadmapping (intuitionlabs.ai). For a DDM program, the appropriate consulting contribution is to help operationalize sponsor-controlled traceability, integration, and governance, not to substitute implementation work for endpoint evidence.

Frequently Asked Questions (FAQs)

What is a digitally derived measure in an FDA-regulated clinical trial?

A DDM is a measure derived from data collected by a DHT. It is not synonymous with the device, raw sensor data, or endpoint. Depending on what it measures and how it is used, it may be a COA, biomarker, component of a multicomponent endpoint, or supporting measurement.

What validation does FDA expect for digital biomarkers?

The evidence should be fit for the context of use. The core ladder is DHT verification, DDM analytical validation, clinical validation, and usability evidence, with explicit analysis of operational errors. The burden rises as the measure moves from exploratory to secondary to primary or label-supporting use.

How should a sensor-based endpoint be validated?

Start with a meaningful concept and target population, not a device. Verify source-signal performance, compare the locked algorithm with an appropriate reference under intended conditions, validate the clinical interpretation in the target population, test intended-user performance, and prespecify acceptance and missing-data rules. The analytical protocol should mirror intended use (navigator.dimesociety.org).

Can an algorithm be updated during a trial?

Yes, but not casually. The sponsor should assess whether the update can alter source capture, processing, output distribution, missingness, or interpretation. A material change requires documented testing and, where appropriate, bridging or revalidation before data from versions are pooled.

Is there an acceptable FDA missing-data rate for DHT endpoints?

No universal percentage applies. Acceptability depends on endpoint role, estimand, duration, mechanism, distribution across participants and treatment groups, and sensitivity of the conclusion. Published technical-error missingness has varied from **6% to 14%** even within one neurological review (pmc.ncbi.nlm.nih.gov). Teams should prespecify validity denominators and conduct sensitivity analyses instead of adopting a generic cutoff.

Does a DDM need formal qualification?

Not necessarily. A sponsor can seek feedback for use in one development program, while a qualification pathway is intended to support reuse within a defined context. EMA's pathway can produce qualification advice or an opinion (ema.europa.eu). The relevant FDA pathway depends on whether the tool is a COA, biomarker, novel drug development tool, DHT method, or product-specific endpoint proposal.

Conclusion

FDA's 2026 paper makes the central decision rule clear: **a digitally derived measure is credible only for its specified context and endpoint role**. The evidence argument begins with what matters to patients, then proceeds through target population, concept, context of use, verified signal capture, analytically validated

computation, clinically valid interpretation, usable operation, and controlled trial execution.

An exploratory endpoint can legitimately be used to learn. A secondary endpoint needs stronger prespecification and interpretation. A primary or label-supporting endpoint must carry a prospective, decision-grade evidence package with locked definitions, acceptance criteria, missing-data strategy, and change control. Moving between these roles is not a matter of adding confidence language. It requires closing specific evidence gaps.

The practical deliverable is a living traceability matrix linked to versioned technical and clinical artifacts. It should expose unsupported jumps, define owners, preserve raw data and provenance, distinguish missingness mechanisms, and trigger bridging when a material component changes. Vendor documentation can populate the package, but the sponsor retains responsibility for ensuring that evidence matches the trial's population, conditions, estimand, and claim.

The most defensible programs will therefore resist two shortcuts: treating device novelty as endpoint value, and treating one validation study as universal permission. Fit-for-purpose evidence is contextual. Endpoint hierarchy determines how much evidence is enough, and prospective traceability shows why it is enough.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://www.fda.gov/medical-devices/digital-health-center-excellence/key-considerations-development-and-use-digitally-derived-measures-clinical-investigations-fda-paper
2. https://www.nature.com/articles/s41746-020-0260-4
3. https://www.fda.gov/media/194348/download
4. https://ctti-clinicaltrials.org/wp-content/uploads/2022/03/CTTI-Digital-Health-Trials-Novel-Endpoint-Acceptance-Recommendations.pdf
5. https://link.springer.com/article/10.1186/s12916-020-01773-w
6. https://www.jmir.org/2022/4/e33537
7. https://navigator.dimesociety.org/your-validation-strategy/4-3-analytical-validation/
8. https://www.nature.com/articles/s41746-025-01513-5
9. https://www.fda.gov/news-events/fda-voices/fda-weighs-development-and-use-digitally-derived-measures-clinical-investigations
10. https://www.ncbi.nlm.nih.gov/books/NBK338448/
11. https://intuitionlabs.ai/
12. https://intuitionlabs.ai/articles/fda-digital-health-technology-guidance-requirements
13. https://www.fda.gov/drugs/development-approval-process-drugs/patient-focused-drug-development-glossary
14. https://www.fda.gov/drugs/biomarker-qualification-program/about-biomarkers-and-qualification
15. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-novel-methodologies-drug-development-guidance-applicants_en.pdf
16. https://www.ema.europa.eu/en/human-regulatory-overview/research-development/scientific-advice-protocol-assistance/qualification-novel-methodologies-medicine-development

17. https://www.nature.com/articles/s41746-024-01322-2
18. https://www.fda.gov/science-research/artificial-intelligence-and-medical-products/fda-digital-health-and-artificial-intelligence-glossary-educational-resource
19. https://pages.nist.gov/nccoe-devsecops/notational-reference-model.html
20. https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf
21. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda
22. https://clsi.org/shop/standards/ep09/
23. https://committee.iso.org/cms/live/live/en/sites/isoorg/contents/data/standard/08/39/83968.html?browse=tc
24. https://www.w3.org/TR/prov-overview/
25. https://intuitionlabs.ai/articles/changing-ai-models-validated-workflows-regression-testing
26. https://intuitionlabs.ai/articles/gamp-5-computerized-system-validation-pharma
27. https://dimesociety.org/digital-measures-physical-activity/clinical-researchers/
28. https://www.nature.com/articles/s41746-020-0259-x
29. https://dimesociety.org/library-of-digital-endpoints/
30. https://www.nature.com/articles/s41746-019-0125-x
31. https://pmc.ncbi.nlm.nih.gov/articles/PMC6060770/
32. https://pubmed.ncbi.nlm.nih.gov/39546781/?dopt=Abstract
33. https://www.nature.com/articles/s43856-025-01137-6.pdf
34. https://dimesociety.org/analytical-validation-library/
35. https://journals.sagepub.com/doi/10.3233/SHTI250868
36. https://www.commondataelements.ninds.nih.gov/sites/nindscde/files/Doc/PD/F3012_Best_Practices_for_Digital_Health_Outcomes.pdf
37. https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2021.736418/full
38. https://academic.oup.com/biomedgerontology/article/78/5/802/6507351
39. https://mobilise-d.eu/work-packages/
40. https://www.fda.gov/drugs/innovative-science-and-technology-approaches-new-drugs-istand-program/innovative-science-and-technology-approaches-new-drugs-istand-program-submission-process

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai