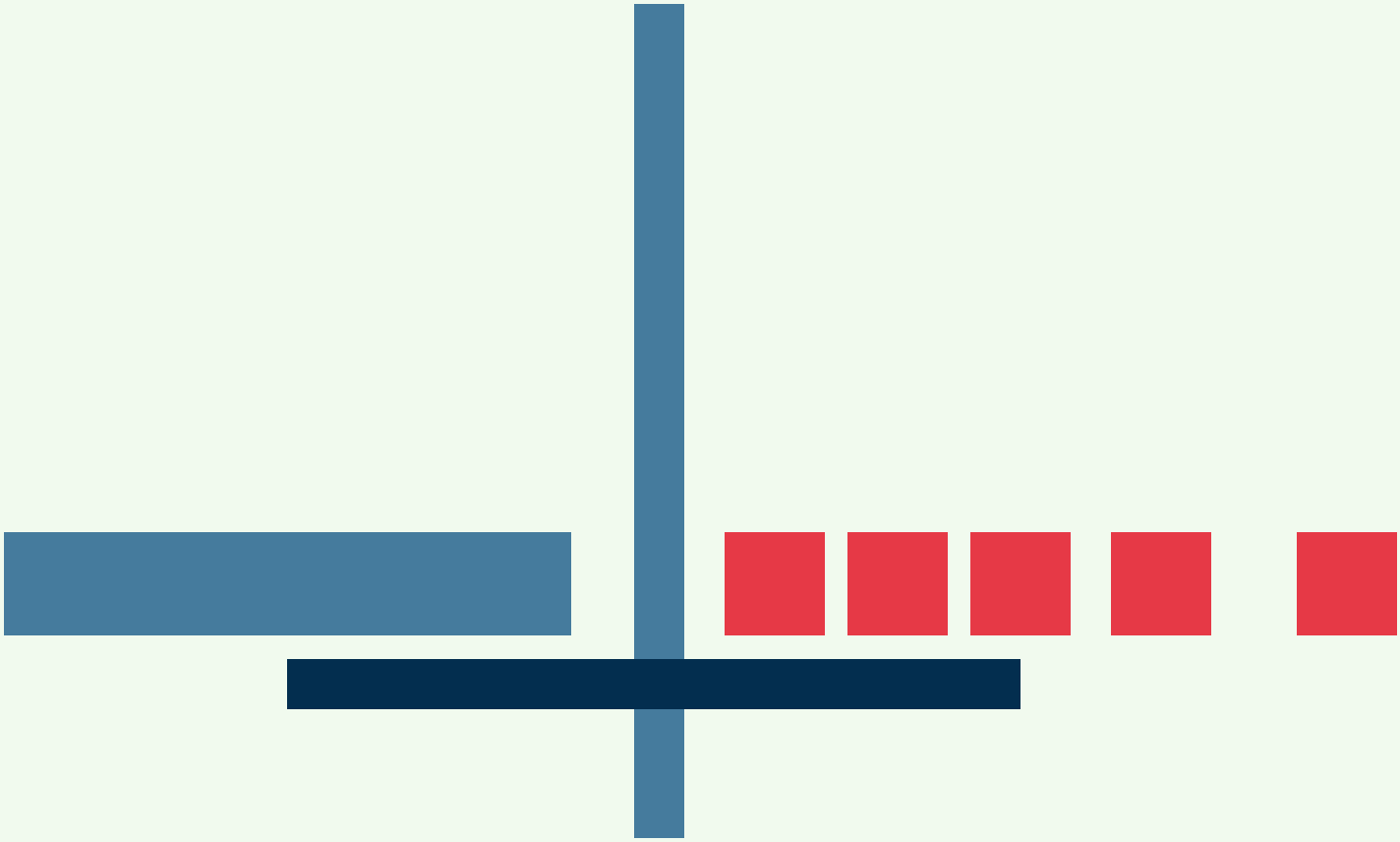


Digital neuroimaging biomarker readiness map: From imaging-derived signals to trial-ready endpoints in CNS drug development

Luciana Bonnot, PhD
Scientific consultant – neuroimaging biomarkers and digital medical device evidence strategy
LuMMiens Consulting | France

Version 1.0 | January 2026

Contents

Executive Summary

Neuroimaging-derived digital biomarkers (NdDB) can strengthen CNS drug development, yet many programs stall before these measures become trial-ready. The issue is rarely the biomarker signal itself. More often, it is the end-to-end measurement chain and its governance: digital requirements sit atop existing trial oversight, blurring roles, evidence expectations, and the level of control needed once outputs begin to influence trial decisions.

Here, a digitally derived endpoint or biomarker is computed by software from clinical imaging acquired in routine radiology workflows. Digital biomarkers may be obtained from various medical devices, including imaging acquisition hardware. They can also be generated, computed, or interpreted by software as a medical device (SaMD), which has a medical intended purpose and performs its function independently of dedicated hardware. In such cases, the SaMD itself is a regulated medical device subject to the corresponding regulatory and quality obligations.

This white paper provides a practical readiness map for NdDB to help pharmaceutical teams make early, defensible decisions before protocol commitments, vendor lock-in, or costly trial amendments. The map begins by sorting candidate biomarkers and then guides teams to the appropriate pipeline maturity level. It also clarifies the role of AI in simple terms and outlines the additional considerations it introduces for trial implementation.

1) Digital biomarkers are constructs.	Regulation applies to the device or software (e.g., SaMD) that generates or uses the biomarker, and to the claims tied to its intended use.
2) Intended use in the trial drives burden.	Moving from exploratory use to stratification or enrichment, or to a primary endpoint, increases expectations for prespecification, traceability, and change control.
3) Pipeline maturity is a deliberate decision aligned with intended use.	The same candidate biomarker can be delivered via research-grade, GCP-grade, or device-grade pathways. Each enables different credible uses.

Trial readiness at a glance

Trial readiness for NdDB requires both fit-for-purpose evidence and control of the imaging/processing chain. Good Clinical Practice (GCP)-aligned operational controls enable traceable, auditable deployment when the biomarker is used as an endpoint or decision driver.

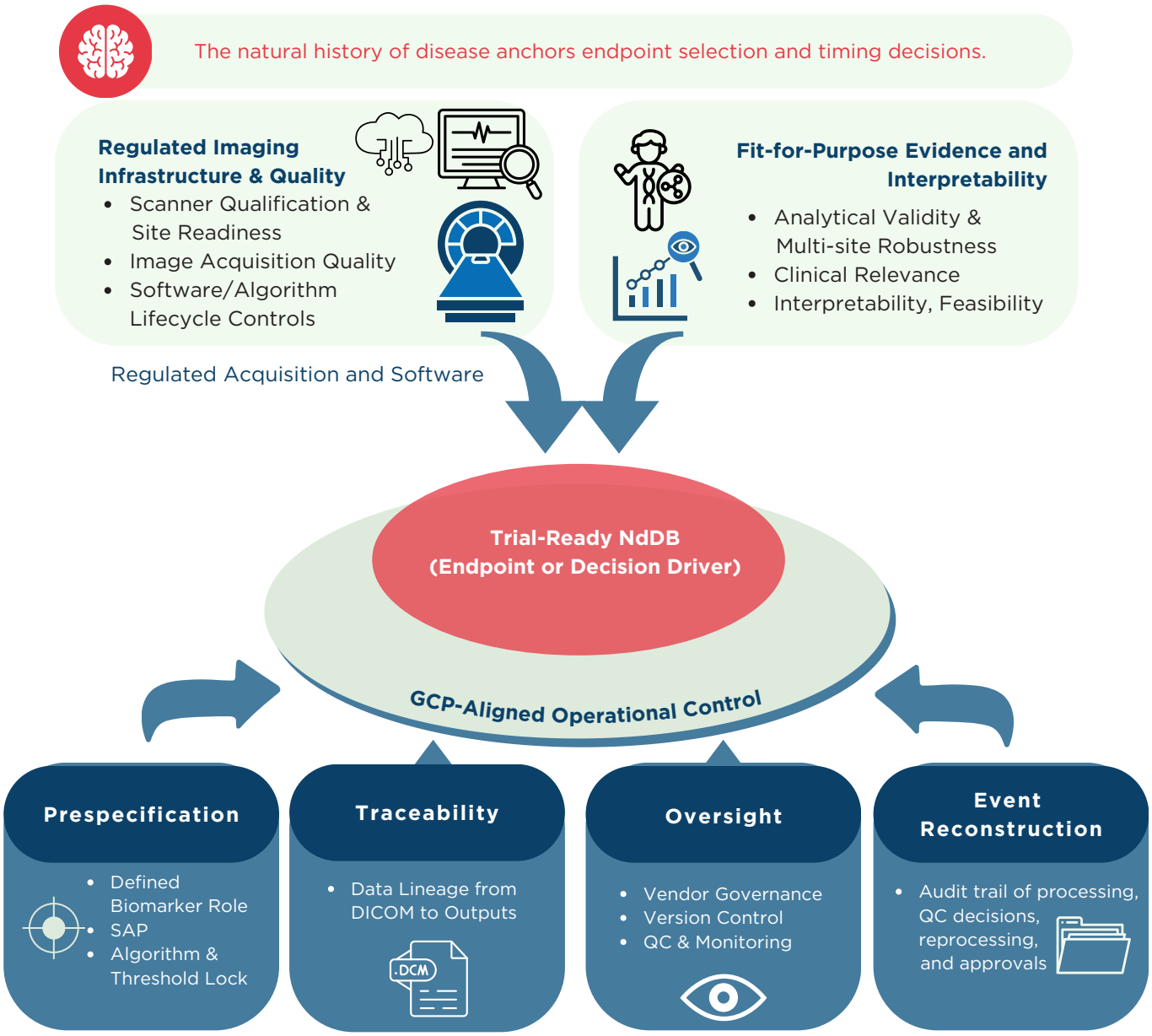


Figure 1. *Trial readiness for NdDB: fit-for-purpose evidence and GCP-aligned controls to ensure traceability, oversight, and event reconstruction from DICOM to outputs. DICOM - Digital Imaging and Communications in Medicine; QC - Quality Control; SAP - Statistical Analysis Plan; GCP - Good Clinical Practice.*

This model applies differently based on your role and the biomarker's intended use in the trial. The next section clarifies who this paper is for and how to use it based on whether you are making biomarker strategy decisions, delivering trial operations, or governing vendors and quality.

Who this is for

This white paper is intended for NdDB strategy owners in pharmaceutical teams who need to decide whether is worth adopting it and which pathway will keep it credible, auditable, and feasible in a clinical trial. The core question for translational medicine, biomarker strategy, and endpoint leadership is:

What is the appropriate clinical anchor for this disease and the context of use, and what is the minimum viable evidence for the next decision gate?

These decision-makers work with cross-functional partners who must execute and defend the pathway. They bring complementary questions on trial risk, imaging operations, site execution, regulatory posture, and vendor control, as shown in Figure 2.

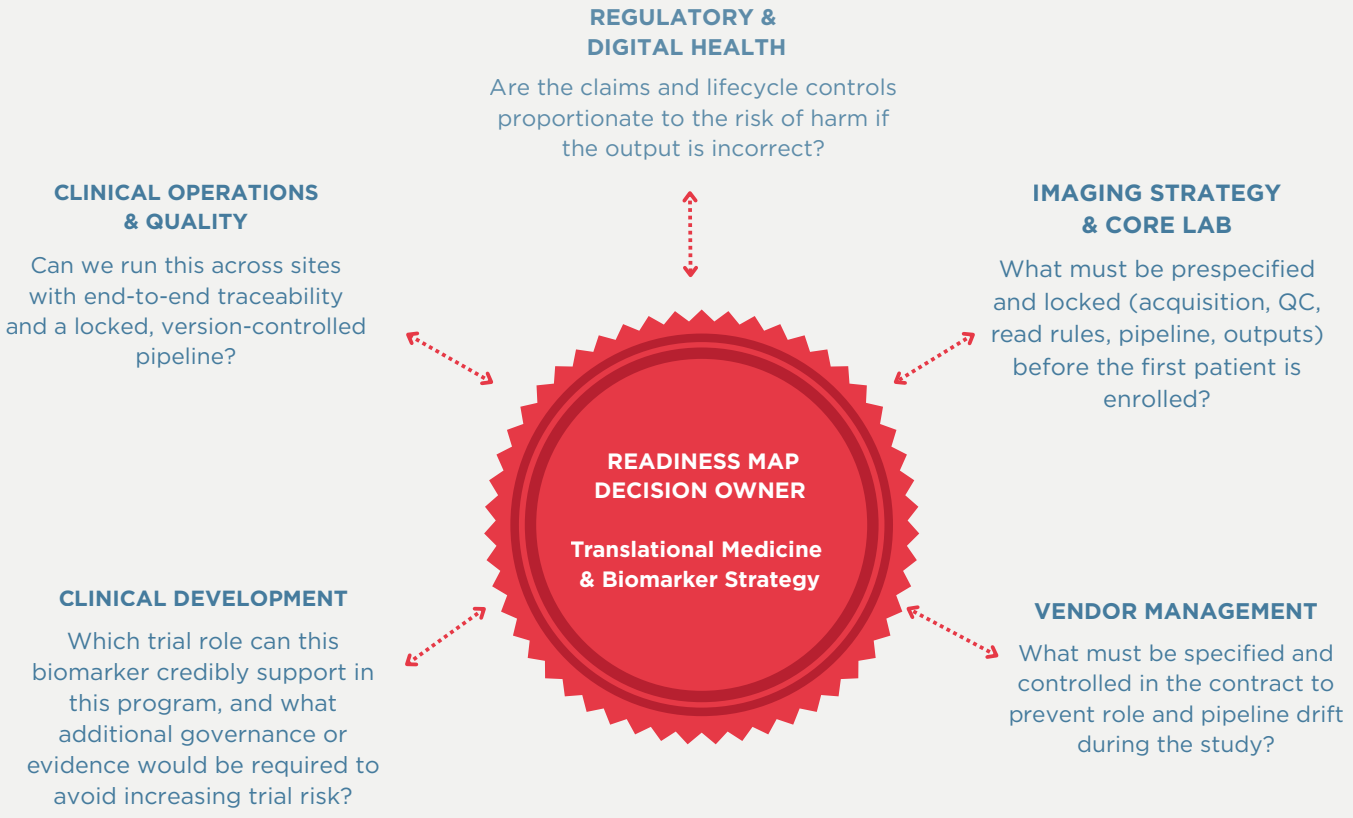


Figure 2. *Typical questions by function that shape biomarker strategy decisions.*

Why adoption fails in practice

In drug development, the key question is not whether advanced image analysis exists. It is whether an imaging-derived endpoint remains credible, auditable, and stable across sites, scanners, populations, and time.

Adoption often fails when constraints on digital medical devices intersect with trial governance, making early role definition more difficult. In digital settings, exploratory imaging measures can more easily move into decision-driving use, since automated outputs are easy to standardize and reuse across teams and workflows. For example, an automated volumetry report may begin as a secondary readout and later be used to support cohort enrichment or stratification because it is generated for every scan and appears consistent across scans. When that shift occurs, the required levels of traceability, version control, and evidence rise sharply, especially in AI-enabled pipelines where updates and performance drift must be controlled. A pipeline that looks robust in a single dataset can break under multi-site variability. A vendor product may be cleared for an intended use that does not match the trial's intended claim.

Core distinctions that prevent category confusion

A digital biomarker is a measurement construct. It is not regulated as a standalone object. Regulatory obligations follow the medical product that generates or uses the measurement and the claims made about its use in context.

Two decisive questions (answer these before anything else):

Is the technology that generates or processes the signal already regulated as a medical device?

What is the intended role of the output in the trial (exploratory, secondary/supportive, monitoring, stratification/enrichment, endpoint, decision support)?

The readiness map: a 3-step decision flow

The map is designed to avoid two common traps: overbuilding governance too early and underbuilding it until it is too late. It uses three steps that can be completed in a structured cross-functional discussion.

1. Triage the candidate biomarker
Clinical establishment, interpretability, novelty, and the clinical variable it must anchor to.

2. Select pipeline maturity grade
Research-grade vs GCP-grade vs cleared/CE-marked software, aligned to the intended trial role.

3. Apply AI as a governance multiplier
Version locking, update policy, monitoring, and change control proportionate to trial criticality.

Triage the candidate biomarker

Before choosing a delivery approach, triage the candidate NdDB. The goal is to avoid a frequent failure mode: selecting a pipeline maturity level before clearly defining what output is being operationalized and what role it can credibly play in a trial. Triage enforces early discipline by separating construct definition, category clarity, and delivery expectations.

Triage is organized along three axes.

Axis 1: Basic eligibility and clarity check (Checklist)

The first step is a checklist-based gate. It ensures that the candidate is sufficiently specified to be discussed as a trial-facing output: the intended clinical variable is defined, the output is unambiguous, the scope of use is explicit, and the candidate is not merely a method, intermediate artifact, or exploratory computation. Candidates who fail this step should not proceed to classification or maturity assessment.

Axis 2: Output classification (Classification router)

Once the candidate passes the checklist, Axis 2 classifies the prespecified output variable using the classification router (Figure 3). This step answers a single question: what kind of output is this?

The router assigns a class (A–H) based on clinical intent, establishment status, and whether the output is framed within radiology reporting practice or represents a different clinical or non-clinical construct. Classification applies per output variable. If a delivery package produces multiple prespecified outputs (for example, a reader judgement plus a quantitative measurement), each output is classified separately.

This step does not assess maturity or regulatory readiness. Its purpose is to provide category clarity by preventing common confusion among reader-based assessments, automated measurements, exploratory metrics, and operational or QC outputs.

Axis 3: Delivery system maturity (Pipeline maturity grades)

After classification, Axis 3 evaluates how the output can be delivered and governed. The maturity grade describes the delivery system, not the biomarker concept itself. As summarized in Table 2, the same classified output may be delivered under different maturity grades, depending on the intended trial role.

Maturity assessment considers traceability, audit readiness, change control, and regulatory posture. It determines which trial roles are credible for the output and the required governance burden. Regulatory clearance or CE marking of a generating tool informs this step, but does not, by itself, validate a trial claim or eliminate the need for fit-for-purpose evidence.

How the axes work together in practice

In practice, triage proceeds in order. Axis 1 confirms that the candidate is a well-defined trial-facing output. Axis 2 clarifies the output. Axis 3 determines how it can realistically be delivered and governed, and which trial roles are plausible. A regulated scanner may anchor image acquisition, but it does not imply that downstream biomarker computation is cleared or CE-marked. Similarly, clinical anchoring affects how quickly an output may reach trial readiness, but maturity grade ultimately constrains adoption.

Together, the three axes prevent premature escalation, align expectations early, and make explicit the trade-offs between ambition, evidence, and governance.

Triage the candidate biomarker

This page shows Axis 1, the checklist-based gate. Candidates who fail this step should not proceed to Axis 2 (classification) or Axis 3 (maturity).

Checklist for Axis 1. Early gate triage checks whether the candidate NdDB is defined and feasible to proceed.

Role and intent
[ ] Intended trial role is defined (endpoint / decision driver / supportive).

Clinical meaning
[ ] Clinical anchor is clear (what clinical variable it represents).
[ ] Available evidence plausibly supports this role in this trial context.

Operational feasibility
[ ] Acquisition is feasible across sites (protocol burden, scanner/protocol variability, expected failure rates).

Computation and specification
[ ] Computation can be prespecified with a controlled lifecycle for this program (versioning, change control, and traceability)
[ ] QC rules and outputs are defined (pass/fail criteria, re-scan/reprocess rules, expected deliverables).

Interpretability and prespecification
[ ] Interpretability is prespecifiable (meaningful change and/or decision thresholds).
[ ] Protocol/SAP impact is clear (what must be locked vs what can remain exploratory).

Decision
[ ] All items satisfied → proceed to Axis 2 (classification) and Axis 3 (maturity/governance).
[ ] Any missing item → pause or stop before investing further.

Pre-check (before Axis 2)
Define the analysis-ready NdDB output (name, units/scale, derivation rules, and accountable sign-off).

Confirm
Does this output unambiguously match the intended clinical variable for the prespecified trial role?

If No → Risk: variable not locked. Teams end up classifying the method or upstream artifacts, leading to misclassification and avoidable rework (redefinition, governance reset, protocol/SAP updates, inconsistent analysis dataset values).

Apply Axis 2 per prespecified output variable. When a delivery package produces multiple prespecified outputs (e.g., reader judgement and a quantitative measurement), classify each output separately and represent the package as a combination of classes (e.g., A+C)

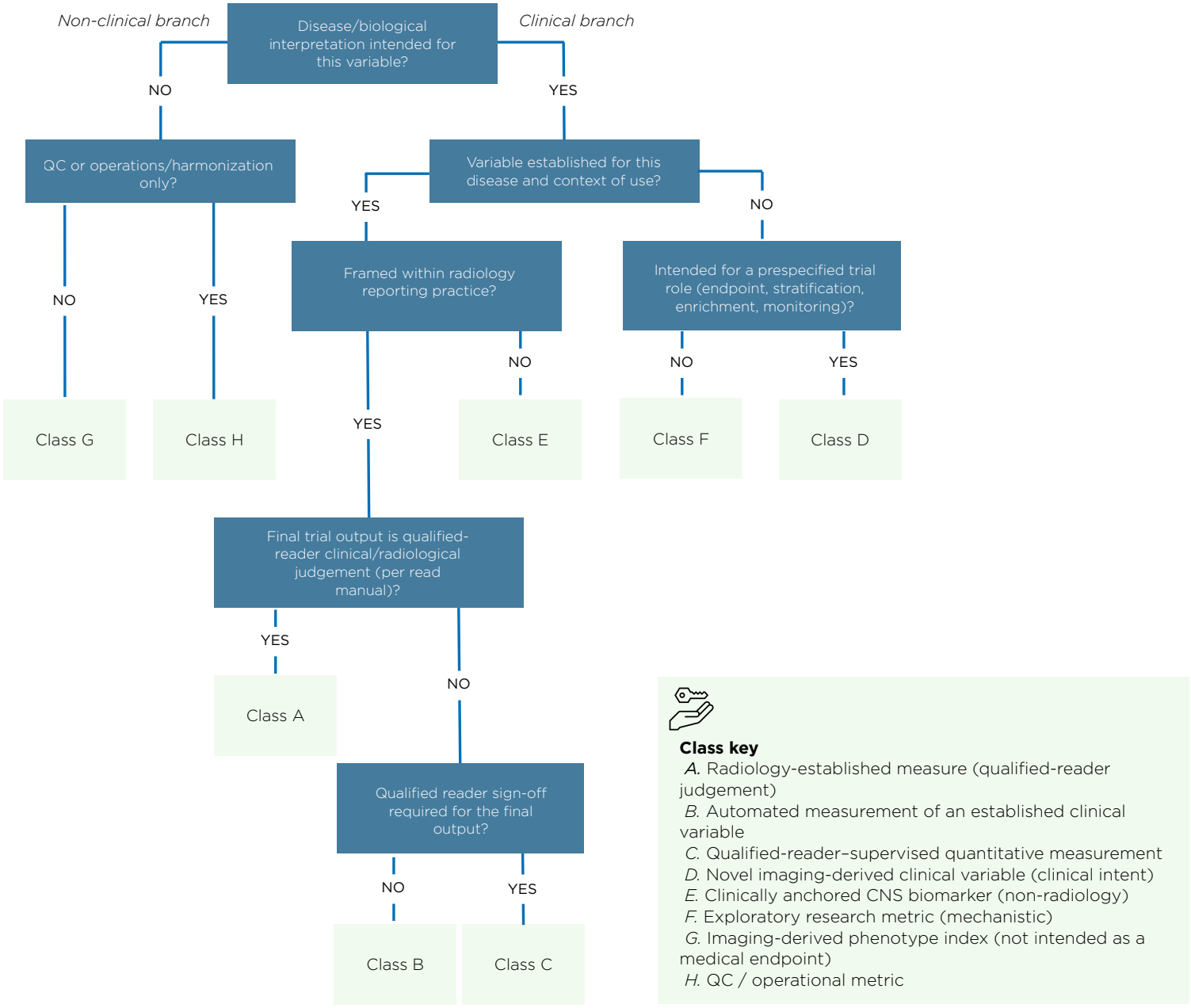


Figure 3. *Axis 2 – Classification router (per prespecified output variable). Routes each prespecified trial-facing output variable to a class (A–H). See Table 4 (on the next page) for class definitions and typical trial uses.*

Table 1. Class definition for the Axis 2 classification router

Class	Figure label	Operational Definition	Typical use in trials
A	Radiology-established measure (qualified-reader judgement)	Clinically established radiology clinical variable expressed as a qualified-reader judgement (categorical/ordinal).	Endpoints or centralized reads when prespecified; adjudication where needed.
B	Automated measurement of an established clinical variable	Algorithm outputs a clinically established variable as the primary value (qualified-reader review optional, not primary).	Supportive or secondary endpoints; decision-driving only when fit-for-purpose, prespecified, and controlled end-to-end.
C	Qualified-reader supervised quantitative measurement	Quantitative measurement of an established clinical variable produced under a prespecified read/measurement manual with qualified-reader sign-off.	Endpoints and centralized measurement workflows; strong traceability, calibration, and inter-reader control.
D	Novel imaging-derived clinical variable (clinical intent)	New clinical variable intended to reflect disease/biology/function.	Exploratory first; evidence-building toward supportive roles, then endpoints only with robust validation.
E	Clinically anchored CNS biomarker (non-radiology)	Clinically established CNS clinical variable not anchored to radiology reporting practice; interpretation anchored to clinical context.	Stratification, enrichment, and target engagement; often combined with other biomarkers and clinical anchors.
F	Exploratory research metric (mechanistic)	Mechanistic or exploratory metric without an agreed clinical variable or prespecified trial decision role.	Exploratory analyses, early-phase insight, and hypothesis generation.
G	Imaging-derived phenotype index (not intended as a medical endpoint)	Pattern/cluster/index derived from imaging for operational or descriptive use, not intended as a medical endpoint.	Cohort characterization, clustering, operational insight; avoid endpoint framing unless re-validated as a medical clinical variable.
H	QC / operational metric	Imaging quality, harmonization, and process indicators (not clinical interpretation).	Site monitoring, harmonization, image quality control, vendor oversight.

Classification applies to the trial-facing output variable. Upstream development artifacts (e.g., feature sets used to derive the output) are out of scope for Axis 2.
Classes describe construct type; deployment burden is determined in Axis 3 based on the intended trial role and delivery-system maturity.

Pipeline maturity grades and fit-for-purpose trial roles

The same biomarker concept can be delivered under different maturity grades. The selected grade sets the minimum delivery controls and constrains which trial roles are credible for a given output.

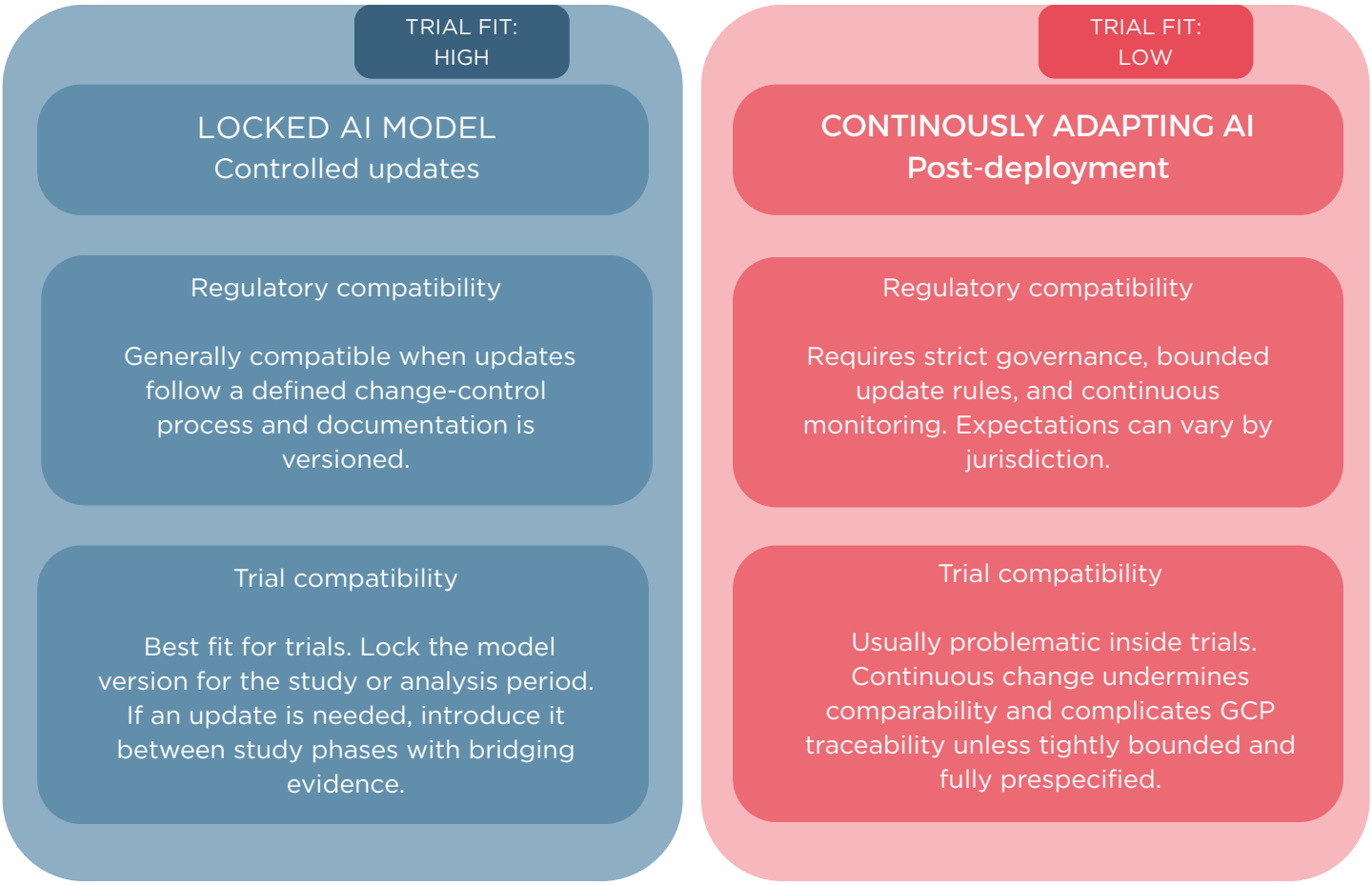
Table 2. Axis 3 - Pipeline maturity grades for NdDB outputs: delivery system expectations, regulatory posture, and credible trial roles

Dimension	Research-grade	GCP-grade (trial measurement system)	Cleared or CE-marked software
Delivery system expectations	Research software applied to clinical images. Flexible workflows. Outputs may evolve during exploration and sensitivity analyses. Traceability and versioning are limited to research practice.	Prespecified outputs for the study, locked versions, centralized QC, and end-to-end traceability from DICOM to reported outputs. Controlled reprocessing rules, defined operational responsibilities, and documented deviations (often via core labs or equivalent services).	Software developed as a medical device with device-grade lifecycle controls. Strong documentation, change control, and operational stability by design, independent of the specific trial claim.
Regulatory posture	Typically not a cleared or CE-marked generating system for the trial claim. Governed as research computation unless intentionally upgraded.	Not necessarily a regulated SaMD product, but treated as a trial-critical computerized system with GCP-aligned controls. Intended use and claims remain trial-defined.	Generating tool is a regulated medical device (e.g., SaMD) for a defined intended use within a jurisdiction. Regulatory status applies to the product, version, and intended use in scope.
Credible trial roles	Exploratory analyses, hypothesis generation, supportive or contextual measures, covariates.	Secondary or supportive endpoints, monitoring, pharmacodynamic readouts, stratification or enrichment only when fully prespecified and controlled end-to-end.	Supportive endpoints, monitoring, decision-driving roles only with fit-for-purpose evidence for the specific trial claim.
Sponsor decision flags	If used for eligibility, stratification, enrichment, or any decision-driving role, governance, validation, traceability, and change control must be upgraded beyond what research tooling typically provides.	Is computation prespecified in the protocol/SAP? Are versions locked for the study? Is there full traceability from source images to final outputs (QC logs, deviations, reprocessing rules)? Are responsibilities and escalation paths defined?	Clearance does not automatically validate the trial claim. Lock the version per study, define bridging evidence for updates, and verify alignment between the device intended use and the trial role.

Note. Maturity levels describe delivery models, not specific products. Regulatory status and intended use depend on the generating tool, its exact version, and the applicable jurisdiction. Grades are not progressive by default; upgrades are triggered by the intended trial role, not by technical sophistication. Maturity grade applies to the delivery system for a prespecified output. Axis 2 defines the output; Axis 3 defines how it is delivered, governed, and made credible for a given trial role.

AI escalation: governance rather than classification

AI does not automatically change risk class, so long as intended use and the significance of information provided do not change. However, it often increases lifecycle risk because performance can vary across populations, settings, and time, especially when updates are frequent.
For clinical trials, stability is as important as accuracy. Endpoints must remain comparable across sites and visits. In practice, this favors locked versions during a study period, with controlled updates introduced between studies.



Practical trial rule: if the digital biomarker output can vary over the life of the study because the model or pipeline is not locked, avoid decision-driving use.

In general, AI-based NdDB implemented as a Software as a Medical Device (SaMD) are strategically important for pharma but introduce specific development challenges. Stable, well-defined features are essential for reproducible biomarkers and generalizable ML models. Increased AI complexity raises uncertainty and lifecycle risk, driving higher evidence and control requirements. The highest regulatory and development burden arises when *complex biomarkers* are coupled *with complex AI modeling layers.*

Sponsor checklist: go/no-go questions

Use this questionnaire as a structured gate before you commit to a trial-critical role. If you cannot answer several items, the main risk is not algorithm performance. The main risk is that the NdDB cannot be executed consistently across sites or defended with the required traceability, version control, and oversight.

Questions to guide the sponsor checklist for NdDB in clinical trials

Gate Area	Question
Role and clinical anchor	Is the intended trial role explicit and prespecified in the protocol and Statistical Analysis Plan?
Role and clinical anchor	Is the clinical anchor defined for this disease, including meaningful change within the trial time window?
Execution and feasibility	Can the biomarker measurement process run consistently across sites and over time?
Execution and feasibility	Is end-to-end traceability feasible from raw images to reported outputs?
Execution and feasibility	Is the operational model defined, including responsibilities, turnaround times, and failure handling?
Updates and control	Can the full pipeline be locked for the study period, including versions, parameters, and reporting rules?
Updates and control	If updates are unavoidable, is change control defined and is there a bridging evidence plan that preserves comparability?
Evidence and robustness	Is the validation plan fit for the intended role, including a defensible reference standard and robustness testing?
Vendors and contracts	Are deliverables, audit access, and strict change control defined in the contract?
Regulatory posture and claims	Is the regulatory posture aligned with the trial claim across jurisdictions in scope?

Common risk patterns and early prevention moves

These are predictable risk patterns that follow from the decision questions in Figure 2 and standard trial governance expectations. They are not exhaustive, but they cover common ways in which NdDB efforts can lose credibility, comparability, or auditability if not addressed early.

Table 3. Risk patterns that arise when Figure 1 decisions are not locked early, and how to prevent them.

Cross-functional trigger question	Risk pattern if not addressed early	Early prevention move
"What trial role does this biomarker currently have, and is it drifting toward a decision-driving use?"	Role creep without a matched upgrade in governance, evidence, and operating controls.	Lock the highest-criticality intended role in the protocol and SAP; match the pipeline maturity grade to that role; define promotion criteria (what evidence and controls are required before using it for decision-driving purposes).
"Can we generate comparable biomarker outputs across sites, scanners, and protocol variations?"	Site and protocol variability dominate signal quality and undermine reproducibility and interpretability.	Define minimum acquisition requirements, allowed protocol ranges, and QC thresholds before site activation; implement centralized QC; document failure handling and controlled reprocessing rules; require evidence of robustness across the expected scanner and site mix.
"Is FDA-cleared or CE-marked status enough for our trial claim?"	Regulatory clearance or CE marking of the product does not, by itself, demonstrate that the biomarker output is validated for your specific trial claim and endpoint role.	Confirm intended use and claims boundary versus your trial role; lock the version for the study; request context-of-use evidence (population, disease spectrum, acquisition context, site variability) and prespecify the output definition in protocol/SAP.
"Can we update the pipeline or AI during the study?"	Moving-target computation undermines interpretability and breaks auditability if changes are not prespecified and bridged.	Adopt a study-level version lock for the algorithm and the whole pipeline; define update windows between studies or analysis periods; prespecify bridging-evidence requirements and amendment triggers; monitor performance while keeping model behavior fixed within the study.
"Can we reproduce values and show lineage from DICOM to outputs?"	Weak data lineage and change control undermine audit readiness and confidence in results.	Treat the pipeline as a trial measurement system: end-to-end lineage from DICOM to outputs, versioned code/configs, QC logs, controlled reprocessing rules, and documented change control thresholds; ensure every output is attributable to a dataset, pipeline version, and QC state.

Where I typically support teams is at the disease-specific decision points behind these risks: defining the clinical anchor and meaningful change, translating intended use into trial-role governance, stress-testing feasibility across real imaging workflows, and designing a defensible evidence and traceability chain.

How to apply the map in your program

This map is general. The critical decisions are disease-specific: clinical anchor, meaningful change, and feasible multi-site imaging operations.

Many teams can use this map to structure internal decisions and align conversations. The main value is making key assumptions explicit early, before protocol design, vendor scope, and site plans are locked.

The most complex decisions are usually disease- and program-specific. They include choosing the right clinical anchor for the NdDB defining what meaningful change looks like within the trial time window, and confirming feasibility against real-world site workflows and acquisition constraints.

When those points are not resolved early, programs often lose months to rework. Teams revisit endpoint definitions, renegotiate vendor scope, and revise imaging manuals to reflect what sites can actually deliver, including acquisition parameters, QC thresholds, and re-scan rules. They may also need protocol amendments once operational constraints surface.

A structured readiness review reduces that risk by aligning stakeholders on a shared definition of the NdDB, its trial role, the required pipeline maturity, and a proportionate evidence and governance plan.

AUTHOR PERSPECTIVE

This readiness map reflects work at the intersection of neurology, psychiatry, and neuroradiology, where imaging signals must remain clinically meaningful while also meeting trial execution and governance constraints. It is informed by experience coordinating multi-site neuroimaging studies and working with cross-functional teams that include clinical development, imaging operations, core labs, and data science. Across neurology and psychiatry programs, I have worked closely with neuroradiology teams to translate imaging-derived measures into decision-ready outputs that are feasible across sites, traceable end to end, and aligned with the intended role in the trial. The document is alive, and updates will be incorporated.

Selected guidance and standards used

This guide is experience-based and intentionally operational. It does not reproduce regulatory text and does not provide legal advice. The concepts and controls described are aligned with widely used guidance and standards listed below, which informed the framework.

EU medical device framework (CE marking)
European Parliament and Council of the European Union. Regulation (EU) 2017/745 of 5 April 2017 on medical devices (MDR) [Internet]. EUR-Lex; 2017 [cited 2026 Jan 28].
Available from: https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng

EU clinical trials legal framework
European Parliament and Council of the European Union. Regulation (EU) No 536/2014 of 16 April 2014 on clinical trials on medicinal products for human use [Internet]. EUR-Lex; 2014 [cited 2026 Jan 28]. Available from: https://eur-lex.europa.eu/eli/reg/2014/536/oj/eng

GCP and risk-proportionate trial quality (trial readiness anchor)
International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH E6(R2): Guideline for good clinical practice. Step 5 (Revision 2) [Internet]. European Medicines Agency; 2016 [cited 2026 Jan 28]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-good-clinical-practice-e6r2-step-5-revision-2_en.pdf

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guideline for good clinical practice E6(R3). Step 4 final guideline (06 Jan 2025) [Internet]. ICH; 2025 [cited 2026 Jan 28]. Available from:
https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH E8(R1): General considerations for clinical studies. Step 5 [Internet]. European Medicines Agency; 2021 [cited 2026 Jan 28]. Available from: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ich-guideline-e8-r1-general-considerations-clinical-studies_en.pdf

Computerised systems, electronic data, audit trail, and event reconstruction - This document should respond to this question: "Is the system fit for use in the trial, validated/controlled, and able to support auditability and data integrity?" (trial governance and inspection readiness)
European Medicines Agency. Guideline on computerised systems and electronic data in clinical trials [Internet]. EMA; 2023 [cited 2026 Jan 28]. Available from: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-computerised-systems-and-electronic-data-clinical-trials_en.pdf

Biomarker and endpoint terminology backbone
FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2016 [cited 2026 Jan 28]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK326791/pdf/Bookshelf_NBK326791.pdf

Software qualification and classification under MDR - This document should respond to this question: "Is it a medical device software, and what class is it?" (product regulatory status and pathway)
Medical Device Coordination Group (MDCG). Update: MDCG 2019-11 rev.1. Qualification and classification of software under Regulation (EU) 2017/745 and Regulation (EU) 2017/746 (June 2025) [Internet]. European Commission; 2025 [cited 2026 Jan 28]. Available from: https://health.ec.europa.eu/latest-updates/update-mdcg-2019-11-rev1-qualification-and-classification-software-regulation-eu-2017745-and-2025-06-17_en

Selected guidance and standards used

IMDRF backbone for SaMD and AI/ML: definitions (N10), risk framing (N12), clinical evaluation (N41), and, where ML applies, GMLP principles (N88).

IMDRF SaMD and AI/ML lifecycle principles (global alignment)
International Medical Device Regulators Forum (IMDRF). Software as a Medical Device (SaMD): Key definitions (IMDRF/SaMD WG/N10FINAL:2013) [Internet]. IMDRF; 2013 [cited 2026 Jan 28]. Available from: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf

International Medical Device Regulators Forum (IMDRF). Software as a Medical Device: Possible framework for risk categorization and corresponding considerations (IMDRF/SaMD WG/N12FINAL:2014) [Internet]. IMDRF; 2014 [cited 2026 Jan 28]. Available from: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf

International Medical Device Regulators Forum (IMDRF). Software as a Medical Device (SaMD): Application of Quality Management System (IMDRF/SaMD WG/N23FINAL:2015) [Internet]. IMDRF; 2015 [cited 2026 Jan 28]. Available from: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-151002-samd-qms.pdf

International Medical Device Regulators Forum (IMDRF). Software as a Medical Device (SaMD): Clinical evaluation (IMDRF/SaMD WG/N41FINAL:2017) [Internet]. IMDRF; 2017 [cited 2026 Jan 28]. Available from: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-170921-samd-n41-clinical-evaluation_1.pdf

International Medical Device Regulators Forum (IMDRF). Good machine learning practice for medical device development: Guiding principles (IMDRF/AIML WG/N88 FINAL:2025) [Internet]. IMDRF; 2025 [cited 2026 Jan 28]. Available from: https://www.imdrf.org/sites/default/files/202501/IMDRF_AIML%20WG_GMLP_N88%20Final_0.pdf

Imaging endpoints in drug/biologics trials (process standards)
U.S. Food and Drug Administration (FDA). Clinical Trial Imaging Endpoint Process Standards: Guidance for Industry [Internet]. Silver Spring (MD): FDA; 2020 [cited 2026 Jan 28]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-imaging-endpoint-process-standards-guidance-industry

Device-grade in clinical investigation expectations) - Standard text is paywalled; reference links to the official ISO catalog entry.
International Organization for Standardization. ISO 14155:2020. Clinical investigation of medical devices for human subjects — Good clinical practice [Internet]. ISO; 2020 [cited 2026 Jan 28]. Available from: https://www.iso.org/standard/71690.html

Selected scientific articles

Koller C, Blanchard M, Hügle T. Navigating through regulatory frameworks for digital therapeutics and biomarkers. Health Informatics J. 2025 Oct-Dec;31(4):14604582251387656. doi:10.1177/14604582251387656.

O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol. 2017 Mar;14(3):169-186. doi:10.1038/nrclinonc.2016.162.