Qity

A PLAYBOOK FOR ML-ENABLED MEDICAL DEVICES

Risk management for machine learning in medical devices

ISO/TS 24971-2 was published in Jun 2026, and rather than replacing the risk management process of ISO 14971 or adding requirements to it, the document directs that process towards the sources of harm that exist because a device learns from data. This playbook works through what it adds, clause by clause, and connects each clause to the life-cycle architecture set out in the companion paper below and implemented in the Qity AI Management System, where the AI System, its models, its datasets and its risks are held as controlled records in the same iQMS as the risk management file.

AUG 2026 EDITION

ISO/TS 24971-2:2026 · QITY.BE

Guidance

ISO/TS 24971-2:2026 is a Technical Specification to be used with ISO 14971:2019 and ISO/TR 24971, and instead of introducing a new process or new requirements it directs the process already in place towards the causes of harm that arise when behaviour depends on data, statistical learning and deployment context. Its scope is the machine learning enabled medical device, referred to throughout as MLMD, and it excludes devices that use large language models or generative AI, so a governance claim built on this document alone does not cover generative behaviour.

> *"One post-market signal may challenge a model, deployment context, risk control and release assumption simultaneously."*
>
> A Life-Cycle Approach to AI Governance in Medical Device Development, 2026

WHY AN MLMD IS NOT ORDINARY SOFTWARE

Three dependencies distinguish an MLMD, and each moves part of the safety argument out of the code and into records the code does not contain.

LEARNING FROM DATA

The model is an artefact of its data

Because an ML algorithm derives the model parameters from training data, device behaviour is a property of the data, the labelling and the separation between training and test data, and can be reproduced only where those versions are known. A device can therefore drift out of specification while its code is untouched.

LEVEL OF AUTONOMY

Oversight has to be designed

Where an MLMD generates options, selects one and executes it with little or no user action, the user is no longer the barrier that catches an incorrect output, so the intervention mechanism and the hand-off point become design decisions. Autonomy can improve patient management and reduce situational awareness at the same time.

EXPLAINABILITY

Opacity is a risk, not a footnote

A clinician who cannot see which factors produced an output cannot judge when to distrust it, so limited explainability creates use-related risk in the workflow. Where those factors can be stated they belong in the information for safety; where they cannot, the document expects compensating risk control measures and a recorded justification.

TWO DEFINITIONS OF RISK

ISO/IEC 23894 and ISO 31000 define risk as the effect of uncertainty on objectives, while ISO 14971 defines it as the combination of the probability of occurrence of harm and its severity, so neither can share a record with the other.

"These perspectives overlap whenever an AI characteristic can contribute to harm, but neither contains the other."

SET ACCEPTANCE CRITERIA FIRST

Establish MLMD acceptance criteria during concept development, before the ML model is tested, because criteria written afterwards are shaped by the results they are meant to judge. They are not criteria for risk acceptability.

Seven places where machine learning changes the work

Clauses 4 to 7 of ISO 14971 do not change, so the table below is the delta against the risk management file already in place. An analysis that stops at "incorrect output" as a software failure never asks *"why statistical performance varies across populations, acquisition devices, sites or time"*, and is *"too coarse to select or monitor the most effective control"*.

CLAUSE	WHAT IT ADDS FOR AN MLMD
4.3 Competence Who performs the tasks	The team shall cover the ML life cycle: good machine learning practice, platform IT and security, algorithm and simulation testing, software validation of the risk control measures, usability engineering to IEC 62366-1, clinical workflow, and data management.
4.4 Risk management plan The one new requirement	Decide during planning whether post-production activities include performance monitoring, updates, or retraining. If monitoring is necessary for safety, the plan **shall** define the methods and processes and the data collected and reviewed. If retraining manages risk, the plan states the criteria for initiating it, how often they are applied, the retraining activities, and the provisions for returning to a previous version.
5.2 Use and misuse What counts as foreseeable	Reasonably foreseeable misuse includes use outside the intended patient population, use with inadequate input data, use error caused by limited transparency, and overreliance on the output. Where performance differs between patient groups, restricting the intended use is permitted.
5.3 and 5.4 Hazards Use the annexes	Annex C adds MLMD-specific questions for identifying characteristics related to safety. Annex B adds examples of hazards, sequences of events, and hazardous situations. Use both with the questions in ISO/TR 24971, Annex A.
5.5 Risk estimation When probability cannot be estimated	Where the probability of occurrence of harm cannot be estimated, estimate the risk on the severity of possible harm alone. Hardware and IT reliability complicate the estimate. Usability evaluation and quality assessment of the training data and test data support a defensible estimate.
7.1 Control options Same order of priority	Inherently safe design first: choice of ML algorithm and ML model, data quality metrics, training data and test data verified for completeness, correctness and consistency, evidence that test data were sequestered, input fields that accept only realistic data, and controlled access to data. Then protective measures: human oversight with an intervention mechanism, cross validation of the output, alarm signals and measures for loss of connectivity. Then information for safety and the hand-off strategy.
7.2 Verification Evidence that the measure works	Test the trained ML model against the acceptance criteria set during concept development. Synthetic data can verify a risk control measure under simulated use conditions. Test a demographic bias control by resubmitting a patient record with one attribute changed. Verify return to a previous version by simulation, and confirm that no new risk is introduced and no existing risk is increased.

WHERE NOTHING CHANGES

For management responsibilities (4.2), the risk management file (4.5), risk evaluation (6), residual risk and benefit-risk analysis (7.3 to 7.6), and the risk management review (9), the document adds no MLMD-specific guidance. ISO/TR 24971 applies.

What to weigh at release, and what to watch after it

CLAUSE 8.1 · OVERALL RESIDUAL RISK

Four factors are added to the evaluation of overall residual risk.

Silent failure. Assume the MLMD can fail without informing the user, and include that case in the evaluation.

Level of autonomy. Autonomy can improve patient management and can reduce the situational awareness of the user. Annex D and IEC/TR 60601-4-1 describe the considerations.

Overreliance. Evaluate use-related residual risk for unwanted bias, including transparency, understandability, and over-trust.

Novelty. Where the MLMD changes current medical practice, analyze that degree of novelty as part of the evaluation.

> Record the acceptable range of MLMD performance here. It is the baseline for post-production monitoring.

CLAUSE 8.2 · WHAT TO DISCLOSE

Disclose significant residual risks as before, with two additions. Write both for the intended user. A usability engineering process can evaluate both.

Transparency

Performance specifications and limitations, including differences in performance between patient groups that result from the selection of training data. State the checks recommended before use.

Explainability

The features that the ML algorithm uses and weights when the ML model parameters are determined, where this can be stated. Where it cannot, record a full justification.

ANNEX A · BIAS

Bias is a systematic difference in how objects, people, or groups are treated. It arises in the data, in the ML model, and in how a user reads the output. Each source needs a risk control measure and verification of its effectiveness. Dataset size is not a risk control measure.

CLAUSE 10 · PRODUCTION AND POST-PRODUCTION

"It directs monitoring and change review towards drift, retraining, site differences, feedback loops and rollback."

01 · COLLECT

Actively collect safety-related information. Consider the level of autonomy, continuous learning, and how the MLMD is used in clinical practice.

02 · REVIEW

Review for drift, meaning gradual and unintended change in MLMD output or behavior, against the performance range recorded at release.

03 · DECIDE

Apply the criteria defined in the risk management plan: improve the concept, retrain the ML model, update the MLMD, or return to a previous version.

04 · ACT

Record the outcome in the risk management file. Retraining and return to a previous version are changes and remain subject to ISO 13485 controls.

The document states what to consider. The records have to stay connected

Every clause above assumes a record that survives change, because a reviewer has to be able to say which ML model version is released, which datasets trained and tested it, and which approval a new post-production signal invalidates. ISO/TS 24971-2 connects ISO 14971 to the causal factors specific to machine learning, but not how those dependencies stay retrievable after several retraining cycles. That is what the companion paper addresses and the Qity AI Management System holds as controlled records.

COMPANION PAPER

A Life-Cycle Approach to AI Governance in Medical Device Development

Read the paper →

Miguel Azevedo, Qity BV, 2026

> *"A document repository can contain all of this information, but only a controlled relationship model makes the dependency query reliable."*

CONTROLLED ENTITIES

The AI System, AI Use Case, AI Model, Dataset, Dataset-Model Association and AI Risk are distinct records, so a change to one shows which decisions it reopens.

TWO-LAYER RISK

ISO/IEC 23894 informs the wider AI risk method while ISO/TS 24971-2 directs ML-specific causal factors into the ISO 14971 analysis, without merging criteria.

CONTROLLED EVOLUTION

Boundaries, methods, evidence and decision authority are defined before change is attempted, and a Predetermined Change Control Plan is assembled from those records.

SIX QUESTIONS TO ASK ABOUT ONE DEPLOYED DEVICE

More than two answers of no indicates that the risk management file describes a device that is no longer released.

CLAUSE	THE QUESTION	STATUS
4.3	Did a person who understands the training data review and approve the risk analysis?	Yes / Partly / No
4.4	Does the risk management plan state the monitoring method, the data collected, and the criteria for retraining?	Yes / Partly / No
7.1	Is there evidence that test data were sequestered from training data for the released version?	Yes / Partly / No
7.2	Has return to a previous version been simulated and checked for new or increased risk?	Yes / Partly / No
8.2	Do the accompanying documents state where performance differs between patient groups?	Yes / Partly / No
10	When drift is detected, is it clear which approvals are reopened?	Yes / Partly / No