



INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

AI Post-Market Surveillance: Locked vs. Continuous Learning

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 · 2026-02-25 · post-market surveillance · ai medical devices · continuous learning · locked algorithms · algorithmic drift · fda regulation · samd · medical device validation · machine learning

Original article: <https://intuitionlabs.ai/articles/post-market-surveillance-ai-locked-continuous-learning>



In this article

1. Executive Summary
 2. Introduction and Background
 3. Regulatory Frameworks for AI/ML-Enabled Medical Devices
 4. Post-Market Surveillance: Concepts and Methods
 5. Continuous Learning AI in Medical Devices
 6. Locked (Static) AI Models in Medical Devices
 7. Comparison of Continuous Learning vs. Locked Models
 8. Post-Market Surveillance Considerations: Data and Evidence
 9. Data Analysis and Evidence-Based Discussion
 10. Implications and Future Directions
 11. Conclusion
-

Executive Summary

Post-market surveillance (PMS) of medical devices is a critical component of ensuring ongoing safety and effectiveness after regulatory clearance. In the rapidly evolving field of artificial intelligence (AI)–enabled medical devices (AIaMDs), PMS faces unique challenges and opportunities. AIaMDs may be designed as **locked algorithms**, fixed once deployed, or as **continuous learning systems** that adapt and improve with new data over time. This report provides a detailed, evidence-based analysis of PMS for AI medical devices, with a focus on the contrast between continuous learning and locked models. Our executive summary highlights key findings:

- **Definitions and Scope:** We define *post-market surveillance* as the activities by which manufacturers and regulators collect and analyze real-world data on device performance after market entry (www.who.int). *Locked AI models* are algorithms “in a locked state when changes are not permitted” (innolitics.com), while *continuous learning AI* refers to systems that retrain or update with each new data point during operation (innolitics.com).
- **Regulatory Context:** Global regulators (FDA, EU, IMDRF, WHO, etc.) recognize the novelty of adaptive AI and are developing frameworks. FDA maintains a public resource of AI-enabled medical devices with marketing authorization ([FDA overview](#)). Its January 2025 lifecycle guidance remains a nonbinding draft proposing recommendations for device design, development and implementation ([FDA draft guidance](#)). EU change control requires assessment of the proposed change: MDR Annex IX section 4.10 requires notified-body approval where changes could affect safety, performance or prescribed conditions of use ([MDR](#)). MDCG 2025-6, question 30, explains that learning changes predetermined by the manufacturer, assessed at initial conformity assessment and documented as specified are not substantial modifications under the AI Act and should not be treated as changes to the certified device under the cited MDR/IVDR provisions ([MDCG 2025-6](#)). FDA’s final AI-device *predetermined change control plan* (PCCP) guidance was originally issued on 4 December 2024 and revised on 18 August 2025. Its

components are the **Description of Modifications**, **Modification Protocol** and **Impact Assessment**. Specified changes implemented consistently with an authorized PCCP do not require another marketing submission ([FDA final guidance](#)).

- **Opportunities and Risks of Continuous Learning:** Continuous learning AI can adapt to new patient data, potentially improving accuracy and personalization over time ([www.ncbi.nlm.nih.gov](#)). However, this brings new risks: *catastrophic forgetting* (where new data overrides previous knowledge), overfitting to non-representative cases, and generalization failures ([www.ncbi.nlm.nih.gov](#)). Surveillance must therefore include ongoing evaluation of effectiveness, safety, and bias as the model evolves ([www.ncbi.nlm.nih.gov](#)) ([pmc.ncbi.nlm.nih.gov](#)). The unpredictability of continuous updates raises questions of “when and how” to re-evaluate a device’s performance ([www.ncbi.nlm.nih.gov](#)). Regulators and developers emphasize the need for rigorous *cyclic revalidation*, robust monitoring plans, and transparent reporting.
- **Challenges of Locked Models:** Locked AI devices are simpler to validate premarket but can become **outdated**. Their performance may “drift” if underlying healthcare practices or patient populations change ([www.ncbi.nlm.nih.gov](#)). Locked algorithms require periodic re-evaluation and updates (often via new regulatory submissions) to maintain safety ([www.ncbi.nlm.nih.gov](#)).
- **Post-Market Surveillance Methods:** Traditional PMS tools (adverse event reporting, registries, and device tracking) apply to AIaMDs, but must be expanded. Methods such as **active performance monitoring** (collecting key performance metrics during routine use), real-world data analysis, user feedback loops, and specialized registries are needed ([www.fda.gov](#)) ([pmc.ncbi.nlm.nih.gov](#)). FDA requires device tracking for certain eligible devices when it issues a tracking order ([FDA tracking requirements](#)). WHO guidance stresses systematic data collection and evaluation of real-world use to trigger corrective actions when needed ([www.who.int](#)). For AIaMDs, surveillance should include monitoring for *algorithmic drift*, emergent biases, cybersecurity incidents, and degradation of performance over time.
- **Authorization Example:** FDA granted De Novo classification to IDx-DR (DEN180001) on 11 April 2018 ([FDA classification order](#); [De Novo review](#)). A subsequent version received 510(k) clearance under K203629 ([FDA clearance record](#)). Emerging examples of continuous-learning in practice remain rare, but regulators and industry are actively studying frameworks – for example, FDA’s Total Product Lifecycle (TPLC) approach.
- **Data and Evidence:** FDA distinguishes 510(k) clearance, De Novo classification and premarket approval in its overview of AI-enabled medical devices ([FDA overview](#)). Radiology constitutes the largest share of AI/ML medical devices ([pmc.ncbi.nlm.nih.gov](#)). Clinical interventions can make model predictions appear incorrect on subsequent evaluation ([Ansari et al.](#)).
- **Future Directions:** The field is moving towards **proactive lifecycle governance**. The EU AI Act’s Article 6(1) product-linked high-risk route requires both Annex I legislative coverage and third-party conformity assessment ([Article 6](#)). The Act entered into force on 1 August 2024, with a general application date of 2 August 2026 ([Commission overview](#)); the Commission’s current FAQ gives 2 August 2028 for rules on AI embedded in products, including medical devices ([Commission FAQ](#)). The FDA and global bodies are pushing for “good machine learning practices” (GMLP) and new guidelines

on periodic re-validation. Big emerging themes include use of Real-World Evidence (RWE) to inform updates, advanced surveillance analytics (e.g., machine learning for signal detection (jamanetwork.com)), and harmonized international standards (IMDRF, WHO). Manufacturers are advised to build PMS into device design (e.g. performance logging, human oversight mechanisms) and to engage regulators early on lifecycle plans (www.fda.gov) (www.ncbi.nlm.nih.gov).

In conclusion, effective post-market surveillance of AI medical devices requires balancing innovation with patient safety. Locked and continuous-learning models present different risk profiles: locked models need safeguards against time-dependent degradation (www.ncbi.nlm.nih.gov), while continuous models require rigorous control of unanticipated changes (www.ncbi.nlm.nih.gov). This report explores scientific, regulatory, and real-world perspectives in depth, providing data, expert consensus, and case analysis. It highlights emerging solutions such as algorithm change protocols, real-world performance monitoring plans, and collaborative registries. Robust PMS frameworks will be crucial for the next generation of AI medical devices to ensure they remain safe, effective, and equitable as technology and data evolve.

Introduction and Background

AI-enabled medical devices (AIaMDs) span clinical decision support, diagnostic image analysis, therapy guidance and patient-facing applications. FDA reviews devices through distinct marketing pathways: 510(k) clearance, De Novo classification and premarket approval. Aggregate device populations should be described as marketing authorizations rather than collectively as clearances or approvals ([FDA overview](#)).

With rapid innovation comes a critical need for **post-market surveillance (PMS)**. PMS is defined, for medical devices, as the activities through which manufacturers and regulators collect and evaluate experience gained from devices that have been marketed, in order to identify safety or performance issues (www.who.int). The goal of PMS is to ensure that devices continue to meet safety and performance requirements throughout their life cycle, and to trigger corrective actions if risks arise. The 2021 WHO global guidance on medical device surveillance states:

“Post-market surveillance is a set of activities conducted by manufacturers to collect and evaluate experience gained from medical devices that have been placed on the market, and to identify the need to take any action” (www.who.int).

PMS for traditional medical devices involves collecting data from adverse event reports, device tracking systems, registries, observational studies, and user feedback. Regulatory mandates vary: for example, the EU Medical Device Regulation (MDR 2017/745) requires manufacturers to implement PMS plans, monitoring, and periodic safety update reports, while in the U.S. FDA enforces adverse event reporting (MDR), device tracking, and may require post-approval studies or section 522 surveys for certain devices (www.fda.gov). AIaMDs introduce **new complexities** to PMS. Unlike classical devices, AI algorithms (especially those based on machine learning) can “learn” from data and change their behavior over time. Surveillance needs to distinguish changes in the input distribution (input or covariate shift), changes in the input–outcome relationship (concept drift), and changes in measured performance. A locked model's stored parameters remain fixed unless the model or software is changed; its performance can nevertheless deteriorate as its environment changes ([IMDRF N67](#)).

A key conceptual distinction arises in how an AI/ML is designed to operate post-deployment:

- A **Locked Algorithm** (or *fixed* model) is deployed in a static form. Once the device is on the market, the AI model's parameters remain unchanged until the manufacturer implements a controlled model or software change. In IMDRF terminology, an ML-enabled device is *"in a locked state when changes are not permitted"* (innolitics.com). A locked model may still be updated by the sponsor later (batch retraining), but it is not continually adapting in real time. Whether an update requires another submission depends on the applicable modification assessment and any authorized PCCP ([FDA software-change guidance](#); [FDA PCCP guidance](#)). Internal quality-system change control is distinct from a new marketing submission.
- A **Continuous-Learning AI** system is designed to adapt continuously as it receives new data during normal use. IMDRF defines *"Continuous Learning"* as *"training that leads to change of an MLMD (Machine Learning-enabled Medical Device) with each exposure to data that takes place on an ongoing basis during the operation phase of the MLMD life cycle"* (innolitics.com). Put differently, every new patient case or set of images could be used to retrain or fine-tune the model incrementally (often via online learning algorithms). Continuous learning is also called *"unlocked"* or *"adaptive"* AI. For example, a radiology AI could in principle refine its detection thresholds each time it processes a new scan.

These concepts are distinct from *batch learning*, which lies between them. Batch learning (also termed *"piecemeal"* or *"offline updates"*) involves periodic retraining on a fixed dataset at discrete times (e.g. monthly or annually) separated from normal operation (innolitics.com). Batch learning describes the timing of training, not a requirement to submit every update to a regulator. A primarily locked device can also have specified updates covered by an authorized PCCP ([FDA PCCP guidance](#)).

The choice between locked vs continuous has profound implications for PMS. Continuous learning offers the promise of improving over time with new data, potentially catching new disease patterns or adapting to shifting populations (www.ncbi.nlm.nih.gov). However, it also introduces unpredictability: the algorithm's decision boundaries may shift in complex ways, raising concerns about repeatability, validation, and oversight (www.ncbi.nlm.nih.gov) (www.ncbi.nlm.nih.gov). Locked models, by contrast, offer stability (the device behavior is fixed and typically well-characterized at approval) (innolitics.com), but risk degradation if the real-world environment diverges from the training data (www.ncbi.nlm.nih.gov).

This report examines **post-market surveillance for AI medical devices**, focusing on the tension between continuous learning systems and locked models. We survey regulatory frameworks (U.S., EU, global), technical challenges, PMS methodologies, case studies, and emerging solutions. We emphasize evidence-based analysis, citing authoritative sources: regulatory guidance, consensus papers, medical literature, and real-world studies. The goal is to provide a comprehensive, in-depth resource that identifies best practices, gaps, and future directions for keeping AI medical devices safe and effective over time.

Regulatory Frameworks for AI/ML-Enabled Medical Devices

International Guidance and Definitions

Global regulatory bodies recognize the unique nature of AI/ML-based devices. The **International Medical Device Regulators Forum (IMDRF)** has published harmonized terminology and principles for AI-enabled and machine learning-based devices. In particular, the final IMDRF N67 document dated 6 May 2022 defines key terms: an *ML-enabled Medical Device (MLMD)* is one that “uses machine learning, in part or in whole, to achieve its intended medical purpose.” It explicitly notes that AI-based software can continue “learning and iteration as additional data becomes available” ([IMDRF N67](#)). IMDRF defines:

- **Continuous Learning:** “Training that leads to change of an MLMD with each exposure to data that takes place on an ongoing basis during the operation phase of the MLMD life cycle.” ([IMDRF N67](#)).
- **Batch Learning:** training that changes the model at discrete times, based on fixed data sets ([IMDRF N67](#)).
- **Locked State:** the device is “in a locked state when changes are not permitted” ([IMDRF N67](#)). The guidance notes that “locked device” has been used to mean either one the developer does not intend to modify, or any device that does not perform continuous learning ([IMDRF N67](#)).

These definitions establish a vocabulary for regulators and industry worldwide. IMDRF lists its *Good Machine Learning Practice (GMLP)* guiding principles, N88, as final, with a publication date of 29 January 2025 ([IMDRF N88](#)). These are distinct from work on more detailed change-control guidance. At the national level, regulators have begun issuing AI-specific policies. The FDA and European Commission, for example, have both published or drafted comprehensive guidance for AI devices (see below). Notably, **WHO** and other global health authorities endorse a total-product-life-cycle approach: the WHO’s 2021 device surveillance guidance explicitly ties PMS to ongoing compliance with safety and performance requirements (www.who.int).

United States (FDA) Framework

FDA reviews medical devices through the applicable 510(k), De Novo or PMA pathway. The terms clearance, classification and approval identify different decisions ([FDA overview](#)).

Lifecycle recommendations: FDA’s January 2025 *Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations* remains draft guidance in the agency’s current index ([FDA guidance index](#)). The draft proposes recommendations for how and when sponsors should describe postmarket performance monitoring and management in marketing submissions ([FDA announcement](#)). These recommendations do not create enforceable duties by themselves. The draft lifecycle document is distinct from the final PCCP guidance.

Predetermined Change Control Plans: The AI-device PCCP guidance is final, with its current version issued on 18 August 2025. A PCCP describes specific planned changes and how they will be developed, validated and implemented. Its three interrelated components are the **Description of Modifications**, **Modification Protocol** and **Impact Assessment** ([FDA final PCCP guidance](#)). A plan must be authorized with the device before it is used. Modifications specified and implemented consistently with that authorized PCCP do not

require another marketing submission. The framework accommodates automatic, manual or combined implementation, including infrequent changes to a primarily locked model; it is not permission for unrestricted adaptation.

Change assessment: A locked model's update does not automatically require a new 510(k). Changes outside an authorized PCCP must be assessed under the applicable modification rules; changes that could significantly affect safety or effectiveness can require a new submission ([FDA software-change guidance](#)). Validation and quality-system documentation remain separate responsibilities.

Postmarket obligations: Applicable adverse-event reporting and quality requirements should be distinguished from device-specific orders or authorization conditions. Section 522 surveillance concerns certain class II or III devices subject to surveillance orders, rather than every AI device ([FDA section 522 guidance](#)). Mandatory patient-level device tracking depends on an FDA tracking order for an eligible device; it is not a universal duty to track utilization ([FDA device tracking](#)).

European Union Framework

The Medical Device Regulation (MDR) and In Vitro Diagnostic Regulation (IVDR) apply to qualifying medical-device software. Existing lifecycle, PMS and conformity obligations continue alongside the AI Act ([MDR](#); [MDCG 2025-6](#)).

Conditional high-risk classification: Under the AI Act's Article 6(1), the product-linked route applies where an AI system is a product, or a safety component of a product, covered by legislation listed in Annex I and the product is required to undergo third-party conformity assessment under that legislation. It does not classify every medical AI system as high-risk merely because it is used in healthcare ([AI Act, Article 6](#)). The Act entered into force on 1 August 2024. The Commission's current implementation FAQ gives **2 August 2028** for rules on AI embedded in products, including medical devices. This replaces the original product-related 2027 timetable and is distinct from the general 2026 application date ([Commission implementation FAQ](#)).

Change control: MDR Annex IX, section 4.10, requires notified-body approval for changes to an approved device where they could affect safety, performance or prescribed conditions of use. The notified body determines whether a new conformity assessment or a certificate supplement is appropriate; the regulation does not require a revised certificate for every improvement ([MDR Annex IX](#)).

MDCG 2025-6, question 30, explains that learning changes predetermined by the manufacturer, assessed at initial conformity assessment and included in the relevant technical documentation do not constitute substantial modifications under the AI Act. Such predetermined changes should not be treated as changes to the certified device under MDR Annex IX section 4.10 or IVDR Annex IX section 4.11. The guidance requires those changes to be clearly specified and addressed in change management and technical documentation. This is an interpretation of the enacted framework, not an exemption for unrestricted learning; the guidance also identifies further PCCP guidance as work under development ([MDCG 2025-6, questions 29–30](#)).

Other Jurisdictions and Global Initiatives

Other major regulators are also active. A historical 2022 CADTH report on **Canada** observed that *no* continuously learning AI medical device had yet been approved there, though locked systems exist (www.ncbi.nlm.nih.gov). The report is a historical account, not evidence of current authorization status.

China's NMPA also discusses whole-life-cycle regulation. The NMPA has initiatives to “optimize whole life cycle regulation” for high-end devices (english.nmpa.gov.cn), likely including AI. Globally, IMDRF and regional harmonization bodies aim to develop shared principles (e.g. IMDRF's Good ML Practices and Europe's MDR guidance).

Across jurisdictions, terminology and lifecycle management are important to comparing locked and adaptive devices. IMDRF distinguishes changes to an ML-enabled device from changes to its environment (**IMDRF N67**). Effective surveillance should combine baseline validation, postmarket data collection, risk management of changes and clear communication with users.

Post-Market Surveillance: Concepts and Methods

Goals and Processes of PMS

The fundamental goal of PMS is to **protect patient safety** by ensuring devices remain safe and effective in real-world use. This involves:

- **Data Collection:** Gathering information from diverse sources (spontaneous reports, clinical studies, registries, device use logs, quality complaints).
- **Performance Monitoring:** Measuring device outputs (e.g. diagnostic accuracy, false positive rate) in practice.
- **Risk Management:** Analyzing data to detect new or increasing risks (e.g. degradation, unexpected errors, cybersecurity issues) and taking corrective actions (design changes, software updates, user education, device recalls).

For medical devices, PMS is both a regulatory obligation and an ethical imperative. WHO emphasizes that PMS “ensures devices continue to be safe and well-performing and to ensure actions are undertaken if the risk of continued use ... outweighs the benefit” (www.who.int). Manufacturers must integrate PMS into their Quality Management Systems (e.g. ISO 13485) and Risk Management (ISO 14971) processes. Key regulatory activities include periodic safety update reports (PSURs) or similar documents summarizing real-world data, vigilance (adverse event) reporting, and implementation of corrective & preventive actions (CAPA) when needed (www.who.int).

For **AI/ML-enabled devices**, PMS has added layers:

- **Model Performance Drift:** Locked model parameters remain fixed unless a model or software change is made. Surveillance monitors changes in inputs, patient populations, clinical practice and outcomes, which may affect performance without changing the model itself (**IMDRF N67**). Input/covariate shift concerns the input distribution; concept drift concerns the input–outcome relationship. Performance degradation is an outcome to investigate, rather than proof that weights changed.
- **Data Bias Detection:** In deployment, an AI may encounter patient subgroups underrepresented in training data. Real-world monitoring should include fairness audits (e.g. by demographics) to catch emergent bias or inequity.

- **Software Updates:** PMS feeds back into the development cycle. If a potential issue is detected, investigate its cause before choosing corrective action or retraining. For either locked or adaptive models, assess proposed changes under the applicable modification rules and any authorized PCCP ([FDA change guidance](#); [FDA PCCP guidance](#)).
- **Usage Patterns:** AI-enabled devices often rely on data inputs or user interactions. PMS may involve analyzing how clinicians use the AI tool (e.g. override rates, user feedback). Continuous learning systems could potentially adapt incorrectly if fed systematically biased user inputs; monitoring must watch for such feedback loops.
- **Cybersecurity & Data Integrity:** Since AI systems often rely on data connectivity, PMS must include vigilance against data breaches or model corruption (e.g. data poisoning attacks). Notably, regulations like the EU's AI Act and FDA guidance stress secure development and logging to support post-market audit trails.

As regulators recognize, effective AI PMS requires a *Total Product Life Cycle (TPLC)* mindset: planning for surveillance at design stage, using real-world evidence (RWE) sources continuously, and updating both the product and the PM plan over time (www.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov).

Post-Market Surveillance Data Sources

Key sources of data for PMS of AI devices include:

- **Spontaneous Reporting Systems:** Reports of adverse events (adverse patient outcomes, device malfunctions) collected via the standard MDR (MedWatch in US, vigilance in EU). However, such reports may miss many AI-specific issues, since misdiagnoses or errors may not be recognized as device-related.
- **Device Registries & Cohort Studies:** Organized collection of cases/platform evaluations, especially for high-risk devices. For example, registries of AI-assisted surgery outcomes or screening results. These allow systematic performance tracking.
- **Real-World Data (RWD):** Analysis of real clinical data (electronic health records, insurance claims, imaging archives) to assess AI performance in situ. The FDA's Sentinel project and other RWD networks could incorporate AI-specific queries. Studies have proposed using RWD to compare AI predictions with actual outcomes (though, as discussed later, causality can be tricky (pmc.ncbi.nlm.nih.gov)).
- **Self-Monitoring Tools:** Some AI devices may include built-in performance logging. For example, an AI imaging tool could record case difficulty metrics, prediction confidence over time, or error rates flagged by users. These logs could be aggregated across hospitals for trend analysis.
- **User Feedback Channels:** Many AI products (especially those integrated into clinical workflow) allow users to flag false positives/negatives or provide corrective input. Systematic capture of this feedback (e.g. an "audit trail" of overrides) is valuable data for PMS, particularly for continuous-learning systems that may incorporate feedback.

- **Periodic Evaluation:** Similar to the lifecycle approach in FDA guidance, manufacturers should perform time-interval performance re-assessments. For instance, every 6-12 months run a validation study on new cases (like a prospective or retrospective reader study or software test) to confirm the AI still meets benchmarks.

Gathering and analyzing these data can be resource-intensive. As **Ansari et al. (NEJM AI)** note, one challenge is that *“positive interventions by clinicians can confound the analysis of AI model performance”*: e.g., if AI kicks off earlier treatment that prevents an adverse outcome, traditional metrics may look worse even though the AI helped ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)). This highlights that PMS data must be interpreted carefully, often requiring advanced causal analysis. Withholding model outputs is a proposed research design, not a routine manufacturer monitoring action: it can create ethical concerns and patient risk. Any such study should require clinical equipoise, ethical review and safeguards for care. Recording interventions alone does not resolve confounding; causal approaches also have practical limitations ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)).

Monitoring Model Performance

To maintain safety and effectiveness, continuous tracking of model performance is essential. Key strategies include:

- **Embedding Key Performance Indicators (KPIs):** Manufacturers should define measurable KPIs for the AI device (e.g. sensitivity, false positive rate, calibration error, response time). These metrics should be collected and trended post-market. The FDA suggests a *Performance Monitoring Plan* outlining the KPIs and how they will be tracked (www.fda.gov).
- **Threshold Alerts and “Retrain Triggers”:** Predefined thresholds for acceptable performance can be established. If the model’s accuracy or error rates fall below a trigger level, this would alert the manufacturer. For continuous systems, this threshold could signal the need for a scheduled retraining or algorithm recalibration.
- **Statistical Process Control (SPC):** Techniques from manufacturing QA (like control charts) can be applied to per-batch error rates. For example, if an AI triage tool normally flags 5% of exams as positive, a sudden jump might flag a drift.
- **Adversarial and Edge Case Logging:** Some modern AI devices log cases where the model is uncertain (low confidence) or fails (rare data patterns). Regular auditing of these cases helps assess if the model is encountering new types of data.
- **Periodic Audits:** Independent evaluations by new sets of observers (like radiologist re-reads) can periodically benchmark AI performance, akin to post-approval studies. For example, an institution might have a workflow where a fraction of cases are reviewed manually for QA.

Overall, the plan should encompass both automated data analysis and human oversight. In the FDA’s draft framework, manufacturers are encouraged to present their PMS methods (including any automated monitoring algorithms) in submissions (www.fda.gov).

Reporting and Risk Management

When surveillance identifies a potential issue, reporting and corrective actions follow. For significant problems (e.g. patient harm or device failure), mandatory reports to regulators are required as usual. Unexplained performance declines require investigation to determine the appropriate corrective action, communication and any applicable reporting obligation. For example, if an AI diagnostic tool is found to have systematically missed certain lesion types, the manufacturer might issue a safety notice and patch the algorithm.

For continuous-learning devices, risk management must address the risks associated with adaptation itself. Changes made under a PCCP still require the verification and validation activities and acceptance criteria in the authorized Modification Protocol. A protocol does not create a presumption that every change is safe, and a departure does not automatically require a full review; the applicable modification assessment determines the submission consequence ([FDA final PCCP guidance](#)).

Continuous Learning AI in Medical Devices

Definition and Rationale

A **continuous learning AI** system is one that refines or updates its model on-the-fly as new data are ingested during real-world use. Technically, this might involve online learning algorithms, incremental retraining with incoming data, or periodic automated re-training without human intervention. The key feature is that the model's internal parameters can change on an ongoing basis **after deployment**. In contrast to static (locked) algorithms, continuous learning allows the device to adapt to new trends in patient populations, technology (e.g. imaging hardware upgrades), or unforeseen circumstances (emergent diseases).

The potential benefits of continuous learning are significant. As the literature notes, continuous-learning AI “*is able to improve its predictions and classifications over time*” (www.ncbi.nlm.nih.gov). This could mean better accuracy as more examples accumulate, adaptation to local population specifics, or incremental learning of rare conditions. In the context of medical AI, one often-cited advantage is that a continuously learning algorithm could incorporate real-world outcomes (ground truth) to refine itself. For example, a sepsis prediction model could adjust if it observes that certain lab trends lead to better outcomes with intervention. Likewise, an imaging AI might improve detection of rare tumor subtypes as it sees more cases.

Continuous learning also promises longer device lifespans: instead of a model gradually degrading as practices evolve, an adaptive model could stay *state-of-the-art*. For instance, say an AI ECG tool was developed using older ECG acquisition devices; a continuous version could learn to handle data from new ECG devices automatically. In public health, continuous AI could quickly integrate knowledge about new strains of disease (e.g. flu variants) or changing epidemiology.

Risks and Challenges

However, continuous learning also introduces **new safety and validation challenges** (www.ncbi.nlm.nih.gov) (www.ncbi.nlm.nih.gov). As a Canadian analysis observes, the very capability of adaptation “*introduces a number of risks and challenges*.” Some specific concerns include:

- **Catastrophic Forgetting:** New training data may override the model's existing knowledge, especially if not carefully managed (a phenomenon called catastrophic forgetting) (www.ncbi.nlm.nih.gov). For example, if a continuous-learning algorithm is retrained on a data stream heavily featuring a new condition, it might lose accuracy on older conditions.
- **Generalizability and Unintended Bias:** A continuous model might adapt too closely to current data, hurting its ability to generalize to other settings. If an AI system is deployed in a new region or hospital with different patient demographics, its adaptive learning could accentuate biases if not corrected. The CADTH report warns that continuous AI may “*not perform as expected when learning from real-world data*” if that data is not representative (www.ncbi.nlm.nih.gov).
- **Unpredictability:** A continuously updating model is inherently a *moving target*. Small changes in data feed can gradually shift the model in unforeseen ways. This unpredictability complicates risk assessment: it becomes harder to guarantee that the model still meets the original safety criteria. Regulators and manufacturers stress the need for “*regular and ongoing evaluation*” of continuous AI systems to detect drift and unintended consequences (www.ncbi.nlm.nih.gov).
- **Validation Timing:** A core question is *when* to re-evaluate effectiveness. Traditional validation (as part of a regulatory submission) provides one snapshot. Continuous learning requires defining triggers for re-validation. Should performance be audited monthly, quarterly, or after a certain volume of new data? These policies are not yet standardized.
- **Data Integrity:** Continuous learning relies on incoming data which may not be curated or fully labeled. There is a risk of “garbage in, garbage out” if poor-quality real-world data inadvertently biases the model. Data drift due to changes in measurement devices or software updates can inadvertently degrade the model.
- **Software Quality and Cybersecurity:** The infrastructure supporting continuous learning (data pipelines, retraining servers) introduces new attack surfaces. For instance, an adversary could attempt *data poisoning* by submitting manipulated cases to alter the model. Regulatory guidance (EU AI Act, FDA) emphasizes robust software development and cybersecurity measures in the lifecycle of AI devices.
- **Transparency and Explainability:** Continuous updates complicate transparency. Clinicians using the device need to know if the algorithm changed. FDA's draft guidance and other experts suggest continuous models should log changes and update logs for review. Moreover, patients and providers must trust that the continuous updates do not make the AI behave unpredictably. Lack of transparency could erode trust rapidly.

Because of these issues, continuous learning AI must have **built-in guardrails**. Proposed risk management strategies include:

- Maintaining a “frozen” version for validation purposes, to compare against the updated one.
- Human oversight mechanisms (e.g. requiring clinician sign-off before algorithm changes are released for clinical use).
- Internal quality controls, such as test suites that each updated model must pass.

- Versioning and audit logs to track exactly how the algorithm has changed (so issues can be traced).

Regulatory Approaches to Continuous Learning

FDA's final AI-device PCCP guidance addresses specified modifications to devices that manufacturers intend to change over time. Its scope includes automatic, manual and combined implementation; a primarily locked device may also have an authorized plan for infrequent updates ([FDA final guidance, sections III and VI](#)).

The plan's **Description of Modifications** sets out specific planned changes, its **Modification Protocol** describes development, verification, validation and implementation methods, and its **Impact Assessment** addresses benefits, risks and mitigations. Modifications described in and implemented consistently with an authorized PCCP do not require another marketing submission. Other modifications require assessment under the applicable change rules ([FDA PCCP guidance](#); [FDA software-change guidance](#)).

An authorized PCCP does not establish overall compliance with every device-law obligation, and the guidance does not provide liability-sharing with FDA. Planning, validation, quality controls and applicable postmarket obligations remain relevant.

In the EU, MDCG 2025-6, question 30, explains the conditions for assessed, predetermined learning changes that do not constitute substantial modifications under the AI Act and should not be treated as changes to the certified device under the cited MDR/IVDR provisions. This must be distinguished from unrestricted learning and from the development of more detailed PCCP guidance ([MDCG 2025-6](#)). Product-linked high-risk obligations follow the Commission's current application timetable, rather than a blanket 2026 start for all medical AI ([Commission FAQ](#)).

Implementation Considerations

Technical Implementation: Manufacturers designing continuous learning AI should architect them with monitoring in mind. This often means splitting the product into (a) a static core that is rigorously validated, and (b) an update mechanism that is carefully controlled. The update mechanism might feed back cases that were misclassified (with clinician confirmation) into periodic retraining. Crucially, any automatic retraining should include integration tests (for safety) before deploying updated weights. The model may maintain an internal *confidence threshold* above which it is allowed to adapt.

Ethical/Social: Continuous learning raises questions about informed consent and transparency to patients. If an AI system changes its behavior, should users or clinicians be informed? How should this information be presented? Answers will need alignment with data governance and medical ethics, especially when the model “remembers” patient data for learning.

Economic/Operational: Continuous learning systems can potentially save time by reducing the need for frequent regulatory submissions. However, they may impose additional burdens: maintaining infrastructure, documenting changes, and robust quality assurance. Healthcare providers need to trust that a continuously learning system will not degrade – which may require contractual assurances from manufacturers.

In all, continuous learning AI offers potential clinical benefits but also demands a strong PMS framework. Without rigorous surveillance and control, a continuously changing medical AI could introduce new, poorly-understood risks. Conversely, with proper checks, continuity of learning can help an AI device remain state-of-

the-art. The remainder of this report analyzes how to achieve that balance in practice.

Locked (Static) AI Models in Medical Devices

Definition and Rationale

In contrast to continuous-learning systems, a **locked AI model** is one whose parameters do not change during normal operation. Once approved or cleared, the algorithm is fixed. If future improvements are needed, they are implemented by releasing a new software version (often requiring fresh regulatory review). In this comparison, locked means that model parameters do not change during normal operation. For example, an image analysis AI is trained on development and validation datasets and then “frozen”. Every patient image is processed with the same model. Updates to this model (retrains, tweaks) are batch processes that occur offline.

Locked models are conceptually simpler and easier to understand. Premarket studies characterize performance under their tested conditions; monitoring remains necessary during real-world use. Established regulatory processes for software modifications (like submission of a new 510(k) for changed indications or substantial algorithm changes) already exist. This model mirrors how other software is regulated and often aligns with a medical device’s life cycle documentation (such as a static “Design History File”).

Benefits of Locked Models

The primary advantage of locking the algorithm is **predictability**. Clinicians and regulators know exactly what the device does and how it was validated. There is no uncertainty from unseen training on new data. This brings several specific benefits:

- **Regulatory Simplicity:** Locked devices follow existing change control processes. Proposed updates need assessment under the applicable change rules and any authorized change plan. An authorized PCCP can cover specified updates to a primarily locked model ([FDA final guidance](#)). As the IMDRF notes, a locked device is one “for which the developer does not have an intention of modifying at the present time” ([innolitics.com](#)). This fits squarely into the conventional device framework.
- **Validation and Reproducibility:** The performance metrics (sensitivity, specificity, calibration) measured in premarket studies remain valid in postmarket use, assuming the environment doesn’t change. There is a stable reference for clinicians to trust.
- **Transparency and Explainability:** With a fixed model, any explanations or interpretations based on the model’s coefficients, if provided, remain valid. Under the EU MDR, manufacturers of devices other than custom-made devices must keep technical documentation up to date ([EU MDR, Article 10\(4\)](#)).
- **Patient Trust and Physician Confidence:** Users often find it reassuring that the algorithm will not “silently change its mind.” There is less feeling of unpredictability compared to an always-learning system.

Limitations and Risks of Locked Models

Locked AI models also have notable drawbacks, particularly over the *long term*. The biggest risk is **obsolescence due to data drift**. Healthcare data distributions can change over time (due to new technology, emerging diseases, demographic shifts). A fixed model trained on data from 2015 might not perform optimally in 2025. For instance, a chest-X-ray model might encounter changes in acquisition protocols or patient populations while still receiving chest X-rays within its intended input modality (**IMDRF environmental-change definitions**). This is a hypothetical example of environmental change, not a documented device case.

Key issues include:

- **Model Drift:** Without retraining, the deployed model may gradually lose accuracy if the underlying data patterns shift. This is akin to how weather forecasting models may degrade if climate patterns change—they are static and need re-calibration. The CADTH review explicitly warns that *“Locked AI can become dated, in that the training data may no longer be representative of real-world data and can experience model drift where the AI performance degrades over time.”* (www.ncbi.nlm.nih.gov).
- **Performance Gaps:** If data or practice patterns change, a fixed model may require investigation and corrective action. An image-quality problem should not be assumed to be caused by locked architecture, or assumed to be solved by continuous learning.
- **Batch Retraining Overhead:** A historical study selected eight resubmitted products from a defined sample of 69 devices identified through seven FDA product codes, using a device-selection cutoff of 1 June 2022. It reported an average interval of 348 days from prior authorization to reapplication. This is a selected historical sample and an application interval, not a market-wide count or time to clearance or deployment (**Original study**).
- **Limited Personalization:** Locked models treat all user populations the same (unless separate versions are released). They cannot learn from local data at a given hospital to improve local performance. For example, if a certain hospital has a unique patient demographic (ethnic mix, disease prevalence) different from the training set, the fixed AI will not adapt to that. Continuous models might mitigate this by adapting to local data streams (though with their own risks).
- **Inefficiency in Learning:** Combining global learning and local deployment is difficult. In a locked approach, each institution may retrain models on local data separately, leading to wasted effort and potentially poorer models that never benefit the broader community. Collecting multi-center data does not itself establish that each subsequent update needs another submission; assess the modification and any authorized PCCP (**FDA final guidance**).

Locked models require effective PMS to identify changes in performance and investigate their causes. Manufacturers should plan periodic evaluation and select risk controls appropriate to the finding. Retraining, workflow changes, user education, software changes or other corrective actions should be assessed on their own merits; neither a new submission nor a recall follows automatically from apparent drift (**FDA software-change guidance**).

Surveillance Strategies for Locked Models

Given the limitations, PMS for locked models focuses on *detection of when performance decays and initiating updates*. Recommended strategies include:

- **Regular Re-Validation Studies:** Schedule periodic studies where the AI outputs are compared to current ground truth. For instance, every 1–2 years, perform a retrospective evaluation on a set of recent cases or images from multiple sites. This can catch drift.
- **Quality Metrics Tracking:** Incorporate usage analytics into installations. If the AI is deployed as part of an imaging system, collect statistics like the percentage of images flagged, distribution of prediction confidences, and repeat analysis rates (cases where AI requests a retake). Any significant change over time could indicate drift.
- **User Feedback Mechanisms:** Implement easy ways for clinicians to flag false negatives/positives. For insured markets, even litigation data (malpractice claims against device) could be monitored as a signal.
- **Periodic Label Updates:** As new kinds of cases emerge (e.g. new disease variants, new imaging devices), plan a cycle of updating the training labels. For example, if an imaging AI sees scans from a new scanner model, collect a sample and relearn features to accommodate it.
- **Alert Thresholds:** If continuous performance monitoring (via sampled outputs) shows metrics falling outside acceptable ranges, issue safety notices or temporarily disable the AI pending revision.

In essence, the locked-model approach treats performance drift as a *risk* to be mitigated by careful monitoring and timely product updates. This is labor-intensive but aligns with existing responsibilities of manufacturers. Regulatory bodies will expect documented PMS plans addressing locked-model drift and any planned cycle for re-training.

A key point noted in the literature is that *both continuous and locked systems need continuous vigilance*. Even a locked model that “never learns” can still degrade if its context changes. The CADTH report comments: “Locked AI will also need periodic evaluation and might lead to a situation of needing to distribute updates or make withdrawals across the healthcare system.” (www.ncbi.nlm.nih.gov). Therefore, even “static” AI devices cannot be totally hands-off. The difference concerns whether model parameters change during operation. FDA’s PCCP framework can cover both automatic implementation and infrequent updates to primarily locked models ([FDA final guidance](#)).

Comparison of Continuous Learning vs. Locked Models

To summarize key contrasts, the table below outlines major differences in approach, benefits, and challenges:

Aspect	Continuous Learning AI	Locked (Static) AI
Definition	Model retrains or updates incrementally with each new data exposure during actual operation (innolitics.com). Sometimes called “adaptive” or “unlocked” AI.	Model parameters remain fixed during operation; any updates occur only when manufacturer deploys a new version through formal channels (innolitics.com).

Aspect	Continuous Learning AI	Locked (Static) AI
Validation	Requires evaluation of the planned changes and their effect on device performance. FDA's final guidance recommends a Modification Protocol with verification, validation and acceptance criteria (FDA PCCP guidance).	A fixed baseline provides a reference for subsequent evaluation. Planned updates also require validation and change control.
Regulatory Pathway	FDA can authorize a PCCP for specified controlled updates; implementation consistent with that plan does not require another marketing submission (FDA final guidance). EU change assessment includes the conditions for assessed predetermined learning changes (MDCG 2025-6).	Applicable modification rules determine whether a new submission or conformity assessment is required. Specified infrequent updates can also be covered by an authorized PCCP (FDA final guidance).
Performance Concept	Trade-off: may improve and adapt, but risk unpredictability. Must guard against <i>catastrophic forgetting</i> and ensure continued generalizability (www.ncbi.nlm.nih.gov).	Stable performance relative to original state, but risk <i>drift</i> if environment changes. Could degrade or become outdated if not retrained.
Risk Management	Control retraining, validation, implementation and communication of planned changes; investigate deviations from the authorized plan (FDA final guidance).	Investigate changes in data, use and outcomes; apply appropriate corrective action and change assessment.
Post-Market Surveillance	Emphasizes automated real-world monitoring of performance metrics, bias, and errors (potentially using AI on real data) (pmc.ncbi.nlm.nih.gov). May incorporate clinician feedback in loop.	Uses standard surveillance (MDR reports, user feedback, periodic audits), but should include revalidation studies and planned updates based on triggered signals.
Adaptation to New Data	High potential for adaptation – can learn from new patient data or imaging sources. For example, could adjust to new disease variants or demographics over time.	No automatic adaptation. Must explicitly collect new data and retrain model offline. Updates lag real-world changes.
Update pattern	Controlled automatic modifications within defined limits. Automatic adaptation does not by itself establish that a particular product has authorization for continuous learning.	Parameters remain fixed in operation; periodic manufacturer updates may still be planned and authorized.

The choice between continuous and locked involves a trade-off between **innovation speed vs. control**. Continuous AI can potentially be more responsive to new information, aligning with agile data science workflows. Both approaches need risk-based ongoing surveillance. Update cadence and regulatory submission needs depend on the specific device, the planned change and the authorized framework, rather than on the locked-versus-learning label alone.

Post-Market Surveillance Considerations: Data and Evidence

Empirical data on AI device performance underscores the importance of robust PMS. Several recent surveys and studies highlight gaps and outcomes:

- **Regulatory Evidence:** FDA's device listings identify marketing authorizations; product-specific evidence needs to be assessed separately from the pathway or the size of the listed population ([FDA device overview](#)).
- **Performance Drift in Practice:** Ansari et al. (NEJM AI, 2025) highlight that AI predictive models often face dataset shifts in use. They describe a challenge for continuous performance monitoring: *“confounding medical interventions....modify outcomes, introducing bias into performance assessment”*

([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)). In other words, an AI that prompted a better treatment might appear to fail at predicting an “outcome” because that outcome was prevented. This insight means PMS for AI cannot simply compare predicted vs. observed outcomes; it must account for the human-in-the-loop effect. The authors call for causal modeling approaches to disentangle true model decay from effective interventions ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)).

- **Signal Detection:** The field of pharmacovigilance has begun leveraging AI (e.g. large language models) to sift through vast data for adverse event signals (jamanetwork.com). By analogy, specialized AI tools may assist in AIaMD surveillance – for example, analyzing free-text radiology reports to spot patterns of AI misses. However, these tools also have biases and false positives (jamanetwork.com), so any PMS AI would itself need validation.
- **Study Interpretation:** Report the endpoint, reference standard, population, device version and study protocol with any performance estimate. Image acquisition, exclusions and clinical workflow can affect measured results; avoid generalizing a limited study into current product performance.
- **Quantitative Evidence:** Treat an apparent change in measured performance as a signal to investigate. Clinical interventions triggered by a prediction can alter the outcome being measured, so a decline in observed predictive performance need not mean the underlying model deteriorated ([Ansari et al.](#)). The authors recommend considering causal pathways before using monitoring metrics to drive update decisions.

Collectively, these data illustrate the variety of evidence that PMS must integrate. Surveys and real-world performance studies complement regulatory clearance data. Notably, **expert consensus** acknowledges that AIaMD PMS is essential: the European Society of Radiology explicitly calls for standardized PMS practices and clinician involvement in monitoring ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)) ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)). They point out a startling gap – a 2025 ESR survey found only ~30% of radiologist users were even familiar with PMS requirements ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)) – underscoring the need for education alongside surveillance systems.

Data Analysis and Evidence-Based Discussion

The surveillance methods discussed above inform broader analysis of AI devices. In particular, we highlight some patterns and expert recommendations:

- **Rapid Approval vs. Slow Data Accrual:** The conflation of rapid deployment and slow evidence gathering is a recurring concern. As one review notes, **AI devices often hit the market with limited user training data** ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)). Postmarket data often lag behind. Consequently, PMS plans must prioritize closing evidence gaps as soon as possible. This might involve multi-center registries or even randomized trials post-approval for high-risk AI.
- **Signal of Performance Decay:** Monitoring baseline performance metrics over months/years can reveal downward trends. For example, if an AI screening tool's positivity rate drifts significantly (after adjusting for population changes), that is a red flag. Statistical process control methods from Six Sigma can be applied: control charts with action thresholds can turn continuous performance data into alerts.

- **Clinical Decision Impact:** Some experts argue that PMS for AI should focus on *patient outcomes* rather than raw model metrics. If an AI helps guide a therapy, one should track outcome metrics (e.g. morbidity, mortality) over time to ensure its benefit is sustained ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)). However, this is complex since outcomes are influenced by many factors beyond the AI. This requires sophisticated causal study designs.
- **Comparison Group:** Given conflicting signals (e.g. Ansari's confounding issue), some recommend comparing outcomes with and without AI assistance in the postmarket period. For example, if an institution sometimes uses the AI (for certain surgeons) and not for others, that natural experiment can inform AI effectiveness. This could be part of a PMS study design.
- **Software Metrics as Signals:** Besides clinical outcomes, software-level metrics can serve as proxies. For instance, a consistently rising false alarm rate might indicate deterioration. A malpractice or complaint database tracking AI-assisted cases can be mined for trends. For example, if multiple claims arise due to a specific failure mode of the AI (e.g. missed lesions in a particular organ), this triggers action.
- **Regulator-Facilitated Studies:** Device-specific surveillance orders can address particular safety and effectiveness questions. Section 522 orders must be distinguished from post-approval conditions and voluntary evidence-generation projects ([FDA section 522 guidance](#)). Public access to postmarket commitments and results can support accountability.

Overall, the evidence underscores that **no surveillance approach is foolproof**. Armed only with passive reports, malfunctions might be missed until too late. Combining multiple data sources (active strategizing) is recommended. The ESR recommendations (radiology) advocate that deploying institutions should share data on AI performance (via feedback forms or network) and that vendors should maintain living post-market clinical follow-up mechanisms ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)).

Implications and Future Directions

Harmonizing Regulation and Innovation

The dichotomy of continuous vs locked models illustrates a central policy challenge: how to encourage beneficial model evolution without compromising safety. Regulators are converging on the view that *proactive lifecycle management* is key. The EU AI Act and FDA TPLC initiatives push manufacturers to integrate monitoring and update mechanisms. We anticipate:

- **Global Standards (IMDRF GMLP):** IMDRF's N88 GMLP guiding principles are final, published on 29 January 2025 ([IMDRF final document listing](#)). The principles address monitoring deployed models and managing retraining risks, as well as informing users about the scope and timing of modifications ([Final N88 document](#)). This completed publication should be distinguished from more detailed change-control work.
- **International Databases:** Similar to Eudamed (EU's device database) and FDA's public listings, we might see registries of AI algorithm performances, e.g., aggregate metrics per breakout by population. This could facilitate benchmarking. Ideally, regulators would require anonymized AI performance logs to

aggregate into national-level surveillance.

- **Dynamic Labeling:** Labels (IFUs) for AI devices may become dynamic over time. Rather than static paper inserts, AI device documentation might be updated continuously. Documentation should describe the intended use, known limitations and relevant human oversight. If a model is updated, the label could note version changes on a website.
- **Stakeholder Roles:** Radiologists and other clinicians are increasingly expected by regulators to participate in PMS. The ESR consensus emphasizes radiologist contribution to monitoring safety ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov)). Nursing staff, technologists, or even patients (via apps) might provide user-driven data. Collaboration between vendors and clinicians on performance logs (e.g. via integrated comment fields in software) is a likely trend.

Technological Aids for PMS

The tools used for PMS will themselves evolve, often leveraging AI for AI oversight. Examples:

- **Anomaly Detection on Data Streams:** Machine learning can be applied to detect anomalies in device usage data. For instance, unsupervised methods could spot when an AI's output distribution shifts. LLMs or big data analytics could scan radiology reports to find narrative mentions of AI errors.
- **Benchmarking Suites:** Simulators and challenge datasets could be run periodically. For example, a consortium might maintain a refreshed dataset of diverse cases; AI devices would have to score periodically on it to demonstrate maintained accuracy.
- **Audit Frameworks for ML:** Analogous to financial audit, ML audit protocols are in development (part of GMLP). These would routinely check training data integrity, version control logs, and performance records of devices in field.
- **Real-World Performance Dashboards:** Cloud-based systems could allow manufacturers to upload de-identified output logs from customer installations for centralized analysis. Regulators or third parties (like patient safety organizations) might run independent analytics on this pooled data.

Economic and Social Considerations

An authorized PCCP can avoid additional marketing submissions for specified modifications implemented consistently with the plan, including updates to primarily locked models ([FDA final PCCP guidance](#)). This submission consequence does not establish a particular cost saving or guarantee clinical benefit. However, it also shifts some burdens onto ongoing processes of validation and monitoring. Cost-sharing models may emerge: for instance, payers might fund post-market registries or data collection for AI devices, given the public health interest. Conversely, if an AI device fails in practice, liability issues become complex. Manufacturers may require clinicians to use devices under specific protocols or disclaim that they operate on trial basis until enough PMS data accumulate.

Another social impact is on **device access**. If continuous-learning devices require internet connectivity for updates and monitoring, resource-limited settings may be disadvantaged or disproportionately affected by lapses. Policymakers will need to ensure that PMS requirements do not inadvertently widen global health

disparities. Open-source AI models could be subject to community-based surveillance as well.

Finally, **patient engagement** will matter. Patients should be informed when AI devices are used in their care and possibly consent-aware to their AI-driven data being used for continuous learning. Transparency initiatives (e.g. an AI label stating the version and last update date) may become standard. Patient-reported outcomes might also feed into surveillance (e.g. symptom tracking after AI-guided diagnosis).

Research and Evidence Gaps

Key areas needing further study include:

- **Continuous vs Batch Update Outcomes:** Empirical comparisons of systems that were continuously updated versus those updated in batches would inform best practices. Currently, long-term studies of continuous AI in real practice are scarce. Pilot programs (e.g. in radiology networks) could provide data on which approach better maintains accuracy.
- **Cost-Benefit Analyses of Surveillance:** Determining the optimal intensity of PMS is an open question. Overly burdensome monitoring (many false alarms) wastes resources; too slack leads to missed problems. Research can help quantify trade-offs (e.g., how many cases should a monitoring plan sample to detect a given performance drop with high confidence).
- **Standard Metrics:** The field lacks standardized metrics for many AI tasks. If every AI uses a different threshold (e.g. risk score) or defines “positive” differently, pooling data is hard. Harmonizing metrics across devices would ease surveillance. This is an area of active discussion (e.g. open AI challenge datasets).
- **Ethical Monitoring:** Feedback loops in continuous learning raise ethical issues. Studies are needed on consent, privacy, and bias correction strategies in continuous systems. For example, continuous retraining on sensitive populations must be done with care to avoid stealth discrimination.

Conclusion

Post-market surveillance of AI medical devices is a field in rapid flux, mirroring the rapid evolution of the underlying technology. Our comprehensive review finds that **locked** and **continuous** learning models each pose unique challenges and advantages. Locked models offer stability and clear validation, but risk becoming outdated and require reactive updates. Continuous learning models offer adaptability and potential for ongoing improvement, but demand rigorous oversight to prevent unpredictable failures.

Both models necessitate vigilant PMS. Established principles of surveillance—data collection, risk management and corrective action—apply, with attention to the model, its inputs and its clinical context. FDA's PCCP framework permits specified changes implemented consistently with an authorized plan without another marketing submission ([FDA final PCCP guidance](#)). The separate lifecycle document remains draft recommendations in FDA's guidance index; it should not be described as imposing new lifecycle-data duties ([FDA guidance index](#)).

Key recommendations from this analysis include:

- **Implement comprehensive monitoring plans:** Every AI device should have a documented PMS plan tailored to its update model (continuous vs locked). This plan should specify performance metrics, update triggers, and data sources (e.g. real-world evidence, user logs, registries). For continuous systems, special emphasis is needed on version control and retraining oversight (www.ncbi.nlm.nih.gov).
- **Engage clinicians in surveillance:** As the European radiology community emphasizes, healthcare professionals must be aware of regulations and actively report suspected issues (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov). Training and education on AI device use/PMS is crucial.
- **Leverage real-world data and AI:** Regulators and manufacturers should use advanced analytics (including AI tools) to process the vast data from device usage for signals (jamanetwork.com). Collaboration on shared real-world datasets could accelerate detection of common issues.
- **Balance innovation with caution:** Updates should remain within the applicable authorization and change-control framework. A plan supports specified, validated modifications rather than unrestricted learning.

Looking to the future, AI in healthcare will only deepen. As models become more powerful (e.g. large multimodal models for medical text and images), the PMS landscape must mature accordingly. We anticipate increased integration of AI surveillance into routine device management. Ultimately, the goal stays constant: ensure that these powerful technologies **do no harm** and improve patient outcomes over the long term. Achieving this will require continuing multi-stakeholder collaboration, investment in evidence generation, and robust regulatory-science frameworks.

Sources: Regulatory references include [FDA final PCCP guidance](#), [FDA's current guidance index](#), [IMDRF final definitions](#), [IMDRF final GMLP principles](#) and [MDCG 2025-6](#). Interpret surveillance methods alongside their study-design and clinical limitations ([Ansari et al.](#)).

Source links

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

1. <https://www.who.int/publications/i/item/9789240015319>
2. <https://innolitics.com/articles/imdrf-guidance-machine-learning-enabled-medical-devices-key-terms-and-definitions/>
3. <https://www.fda.gov/medical-devices/digital-health-center-excellence/artificial-intelligence-enabled-medical-devices>
4. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/artificial-intelligence-enabled-device-software-functions-lifecycle-management-and-marketing>
5. <https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng>
6. https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en?filename=mdcg_2025-6_en.pdf
7. <https://www.fda.gov/media/166704/download>
8. <https://www.ncbi.nlm.nih.gov/books/NBK605105>
9. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12380398/>
10. <https://www.fda.gov/news-events/press-announcements/fda-issues-comprehensive-draft-guidance-developers-artificial-intelligence-enabled-medical-devices>

11. <https://www.fda.gov/medical-devices/postmarket-requirements-devices/medical-device-tracking>
12. <https://www.govinfo.gov/content/pkg/FR-2022-01-21/pdf/2022-01147.pdf>
13. https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN180001.pdf
14. https://www.accessdata.fda.gov/cdrh_docs/pdf20/K203629.pdf
15. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11634576/>
16. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
17. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
18. <https://digital-strategy.ec.europa.eu/en/faqs/navigating-ai-act>
19. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2822418>
20. <https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/postmarket-requirements-devices>
21. <https://www.imdrf.org/sites/default/files/2022-05/IMDRF%20AIMD%20WG%20Final%20Document%20N67.pdf>
22. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/deciding-when-submit-510k-software-change-existing-device>
23. <https://www.imdrf.org/documents/good-machine-learning-practice-medical-device-development-guiding-principles>
24. <https://www.fda.gov/medical-devices/digital-health-center-excellence/guidances-digital-health-content>
25. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/postmarket-surveillance-under-section-522-federal-food-drug-and-cosmetic-act>
26. https://english.nmpa.gov.cn/2025-10/14/c_1138492.htm
27. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence>
28. <https://pmc.ncbi.nlm.nih.gov/articles/PMC9994700/>
29. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12701188/>
30. https://www.imdrf.org/sites/default/files/2025-02/IMDRF_AIML%20WG_GMLP_N88%20Final.pdf

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>

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.