Association for Pathology Informatics

3580 Innovation Way, Suite 104, Hermitage, PA 16148
www.pathologyinformatics.org

December 1, 2025

Center for Devices
and Radiological Health
US Food and Drug Administration

RE: Docket No. FDA-2025-N-4203 - Public Comment to FDA on "Measuring and Evaluating AI-Enabled Medical Device Performance in the Real World"

Introduction

The Association for Pathology Informatics (API) thanks the FDA for the opportunity to comment on the "Measuring and Evaluating AI-Enabled Medical Device Performance in the Real World" request. We support the agency's efforts to build robust frameworks to ensure the safety, effectiveness, and reliability of AI systems throughout their lifecycle, especially after deployment in clinical settings.

Below are API's responses to the FDA's questions posed in the docket. These answers draw from our members' experience with pathology, computational diagnostics, and informatics.

1. Performance Metrics and Indicators

- Clinical metrics such as sensitivity, specificity, area under the ROC curve (AUC), calibration, positive predictive value, negative predictive value, false positive and false negative rates, and subgroup error rates are crucial.
- Metrics should be clearly defined and consistently weighted across safety, reliability, consistency, and fairness.
- Real-world clinical use should be evaluated over multiple practice cycles (6–12 months) to account for population drift and workflow adaptation.

2. Real-World Evaluation Methods and Infrastructure

- Prospective monitoring should combine automated detection (prediction vs. ground truth logging) with expert review.
- Hybrid methods that use thresholds to trigger alarms alongside human oversight are recommended.

- Infrastructure should include data pipelines, model version control, audit logs, feedback loops, and monitoring dashboards.

3. Post-market Data Sources and Quality Management

- Sources should include EHRs, pathology and diagnostic logs, outcomes registries, device logs, and user feedback.
- Data standards, harmonization, and validation are critical.
- Clinical outcomes and override rates should inform recalibration and updates.

4. Monitoring Triggers and Response Protocols

- Triggers may include performance drops (e.g., decrease in AUC), shifts in input distributions, or rising subgroup error rates.
- Protocols should define investigation, temporary suspension, retraining, and redeployment steps.
- Remediation plans must include rollback capability and user communication.

5. Human–AI Interaction and User Experience

- Patterns such as override frequency, alert fatigue, and integration into workflow affect performance.
- Training, model transparency, interpretability, and user feedback channels are essential to safe use.

6. Additional Considerations and Best Practices

- Barriers include inconsistent data standards, limited resources, and institutional inertia.
- Incentives and collaborative consortia can support continuous monitoring and improvement.
- Privacy, security, and governance—especially deidentification and auditability—must remain central.

In addition, API would like to highlight these additional critical considerations not explicitly covered in the current questions: legal responsibility, billing/reimbursement, and AI supervision/autonomy.

- **Legal Responsibility/Liability**
  - Shared accountability: Outputs such as chart summaries, diagnostic predictions, or alerts should not be the sole legal responsibility of either the AI vendor or the clinician. FDA should encourage a shared responsibility model where liability is proportioned based on vendor transparency, validation, and
  - monitoring practices, alongside the clinician's duty of care.

  - Safe harbors: FDA could define conditions (e.g., robust monitoring, transparent labeling, clear instructions for use) under which vendors and clinicians have reduced liability exposure.

- Failure attribution: Guidance should clarify how liability is determined in cases of model drift, poor retraining, or ignored warnings.

- **Billing and Reimbursement**

  - Clarify eligibility: AI-based predictions derived from existing data (e.g., sepsis risk scores, treatment response predictions) require clear guidance on whether they constitute billable "tests" or are considered decision support bundled with other services.
  - Criterion of new information: Billing should be permitted only when AI delivers new, actionable clinical insights beyond routine interpretation of existing data.
  - Validation requirements: Separate reimbursement should be contingent upon rigorous validation, ongoing monitoring, and evidence of clinical utility.
  - Equity considerations: FDA should acknowledge the reimbursement implications of AI adoption to avoid widening disparities between resource-rich and resource-limited care settings.

- **Supervision and Autonomy**
  - Autonomous orders: FDA should address whether AI systems may autonomously place lab or imaging orders and under what conditions.
  - Risk stratification: Low-risk, routine orders may be appropriate for AI initiation with audit trails; higher-risk interventions should always require clinician sign-off.
  - Guardrails: All autonomous actions must include human override capabilities, audit logs, and safeguards against overreach.
  - Integration with liability: Any allowance for autonomy must align with the liability framework outlined above.

- **Transparency**
  - Should pathologists be required to indicate in their report when a specific AI tool was used during the diagnostic process?

- **Modularity**
  While it is well understood that using digital solutions and whole-slide images in clinical practice falls under the LDT process, we recommend that the FDA clearance process for these tools—when cleared by the FDA—be refocused on approving individual components rather than fixed combinations.

  This approach would involve clearing each device separately, such as:
  - **Monitor**
  - **Whole-slide image viewing software**
  - **Whole-slide scanner**

Healthcare facilities and providers could then combine these individually cleared components into a validated pipeline, followed by a clearly defined verification step to confirm performance within expected specifications. This would enable the use of cleared device combinations for primary diagnosis without requiring the costly, time-consuming, and restrictive process of clearing only specific pre-defined combinations.

Mapping to FDA's Existing Questions
Q1 Performance Metrics: Include accountability/liability as a measurable dimension.
Q2 Real-World Evaluation: Assess AI-initiated actions and frequency of clinician overrides.
Q3 Data Sources: Incorporate claims/billing data into monitoring frameworks.
Q4 Monitoring Protocols: Define escalation triggers if AI autonomy leads to adverse outcomes.
Q5 Human–AI Interaction: Evaluate the audit burden on clinicians and clarify oversight standards.
Q6 Additional Considerations: Explicitly call out liability, billing, and autonomy as part of the lifecycle framework.

In conclusion, API urges the FDA to expand its AI performance monitoring framework to address liability allocation, billing guidance, and standards for autonomy and supervision. While the current focus on monitoring, drift detection, metrics, and infrastructure is important, explicit guidance on these additional issues is essential to reduce uncertainty, align incentives, and safeguard patients. Without such clarity, AI adoption in medical practice may be slowed by liability risks and misaligned economic incentives.

Sincerely,

Lisa-Jean Clifford

Lisa-Jean Clifford
President, Association for Pathology Informatics