COLLEGE of AMERICAN PATHOLOGISTS

December 1, 2025

The Honorable Martin A. Makary MD, MPH
Commissioner
US Food and Drug Administration (FDA)
10903 New Hampshire Avenue
Silver Spring, MD 20993

Re: Request for Public Comment: Measuring and Evaluating AI-enabled Medical Device Performance in the Real-World [Docket No. FDA-2025-N-4203]

Submitted via Electronic Submission to www.regulations.gov

Dear Commissioner Makary,

The College of American Pathologists (CAP) appreciates the opportunity to comment on the Food and Drug Administration's (FDA) Request for Public Comment *Measuring and Evaluating Artificial Intelligence-enabled Medical Device Performance in the Real-World*. The FDA has noted that the objective of this request for public comment is to obtain comment and feedback on a series of questions related to the current, practical approaches to measuring and evaluating the performance of AI-enabled medical devices in the real-world, including strategies for identifying and managing performance drift, such as detecting changes in input and output. As the world's largest organization of board-certified pathologists and leading provider of laboratory accreditation and proficiency testing programs, the CAP serves patients, pathologists, and the public by fostering and advocating excellence in the practice of pathology and laboratory medicine worldwide. As physicians specializing in the diagnosis of disease through laboratory methods, pathologists deliver high quality diagnostic services to patients and other physicians. The pathologist's diagnosis and value are recognized throughout the care continuum and many patient encounters. For almost 80 years, the CAP has been the advocate for pathologists, patients, and the public when it comes to improving laboratory quality and assuring that patients receive the right test, at the right time, and with the right result.

The CAP is eager to work with the FDA to share our experience with measuring and evaluating AI-enabled medical device implementation and ongoing monitoring in the real-world laboratory setting, since much of what we do to oversee use of conventional assays applies to AI-enabled devices. Pathologists are the physician leaders of laboratories and are responsible for ensuring laboratory services and results meet patient care needs. With responsibility for oversight of clinical and anatomic pathology laboratories, pathologists have a leadership role in the selection, configuration, deployment, application, and monitoring of artificial intelligence (AI) systems involved in the pre-analytical, analytical and post-analytical phases of laboratory workflow. This

College of American Pathologists
1001 G Street, NW, Suite 425W
Washington, DC 20001
202-354-7100

includes evaluating AI-enabled medical devices prior to implementation, verifying their clinical use, and measuring and evaluating AI-enabled medical device ongoing performance. The best performance of AI in pathology accrues when machine capabilities provide pathologists with added information to extend analysis, enhancing rather than replacing human expertise. AI and machine learning systems can be an additional source of information for a physician to integrate into the overall diagnostic report, but these systems cannot provide a medical diagnosis or be responsible for decision-making in patient care. Physician review and approval are necessary. Put differently, current AI applications in pathology are assistive or augmentative—providing additional information to support interpretation, not autonomous diagnosis.[1]

Assessing AI devices prior to laboratory implementation requires transparent, comprehensible information from device developers about the model development data, performance data, performance metrics, performance monitoring, limitations, installation and use, and customizable features much like similar information is provided in the IFU for FDA cleared or approved laboratory tests. In addition, information on verification steps as part of marketing submissions will help pathologists understand how to conduct verification using local data and the role of the local medical director to oversee implementation and monitor performance over time. Information on expectations for local use will also enable pathologists' awareness of all information related to the likelihood, frequency, and requirements of re-verification prior to acquisition of an AI-enabled device. Pathologists should have information on the techniques and authority to measure AI performance and respond to performance problems to ensure the ethical application and clinical validity of AI systems in the laboratory.

Our members have extensive expertise in providing and directing laboratory services under the Clinical Laboratory Improvement Amendments (CLIA) regulations, which require compliance with requirements through a quality system approach for overall operations and administration of the clinical laboratory. The CAP is a federally deemed CLIA laboratory accrediting organization, and the CAP laboratory accreditation checklists help ensure high-quality diagnostic testing. The CAP Accreditation Checklists require the laboratory director to ensure that the performance specifications for new tests, instruments, and methods introduced to the laboratory have been properly validated or verified prior to being used for patient testing, including AI and machine learning algorithms. Consequently, the expansion of pathologists' responsibilities to include AI constituted an important new element in pathologists' role as CLIA-laboratory directors and section directors. CLIA—as well as the CAP laboratory accreditation requirements—specifies quality practices in the laboratory that go beyond operational requirements defined by a manufacturer of a medical device and approved by the FDA. It is essential that pathologists use their experience as CLIA laboratory directors and section directors to ensure that FDA-cleared/approved devices are used appropriately,

[1] American Medical Association. CPT Appendix S: AI taxonomy for medical services & procedures. December 30, 2024. Available at https://www.ama-assn.org/practice-management/cpt/cpt-appendix-s-ai-taxonomy-medical-services-procedures

as the inappropriate implementation of any FDA-cleared device can result in harm to patients.

AI deployment in pathology is in early stages with relatively few AI models deployed in a limited number of laboratories. Answering the technical questions posed by the FDA is challenging. At this point in time, it is difficult to identify generalizable best practices for AI performance monitoring in pathology. We are committed to working with the FDA to evaluate and monitor AI implementation in the laboratory and anticipate and adapt to challenges. In clinical laboratory testing, pathology has the most direct analog to what the FDA is asking about in this request for comment—that is, there are analogies between AI performance monitoring and quality management in clinical laboratory testing more generally. Indeed, AI-enabled medical devices share some features of conventional laboratory testing methods that are addressed by the CAP's accreditation requirements for clinical laboratory testing, which help ensure laboratories implement and maintain systems and processes that are associated with quality. For example, the CAP accreditation requirements note that owing to potential drifts and shifts in performance over time, laboratories should establish controls, metrics, corrective actions, and procedures to address changes to the test system that could affect clinical results; such practices are similarly necessary for laboratories to implement with the use of AI-enabled medical devices.[2]

Consequently, the CAP's experience—which includes our members' experience as CLIA laboratory directors, section directors, and peer reviewers for CAP's Laboratory Accreditation program—is useful for identifying successful strategies and challenges going forward in measuring and evaluating real-world performance of AI-enabled medical devices.

CAP Responses to the FDA Questions

The CAP will respond to selected questions applicable to our members' expertise in laboratory medicine and AI oversight.

1. Performance Metrics and Indicators

a. What metrics or performance indicators do you use to measure the safety, effectiveness, and reliability of AI-enabled medical devices in real-world clinical use?

CAP Response: *Machine learning encompasses a broad range of techniques, from simple linear regression to complex deep neural networks, each with distinct assumptions, behaviors, and levels of interpretability. This diversity*

[2] Furtado LV, Ikemura K, Benkli CY, et al. *General applicability of existing College of American Pathologists accreditation requirements to clinical implementation of machine learning–based methods in molecular oncology testing.* Arch Pathol Lab Med (2025) 149 (4): 319–327.

makes it challenging to define a single, standardized approach for establishing and maintaining AI trustworthiness, as different model architectures may require different validation and monitoring strategies. Consequently, the framework for measuring and evaluating the performance of AI-enabled medical devices in clinical diagnostics should be relevant to the AI model and acknowledge the framework for laboratory oversight. This will ensure that device users have the best information for verification and validity. In addition, pathologists and laboratory professionals will ensure that AI-enabled medical devices meet the same standards of safety, quality, and accountability as conventional tests. Pathologists' engagement in the development, deployment, and configuration of AI technology helps to ensure that the pathologists' unique knowledge of pathology and laboratory medicine, including patient safety issues, risks, workflow, and other challenges, are incorporated into AI systems.

Following FDA clearance of a medical device, pathologists review the manufacturer labeled information to determine whether to implement the device in their laboratory. This review involves verifying that a device works as intended in the local setting. It also involves placing a process around the device for ongoing performance monitoring to detect any shift in performance outside of expected values as soon as possible to ensure patient safety. The CLIA program requires laboratories to conduct these calibration verification procedures to substantiate the continued accuracy of the test system throughout the laboratory's reportable range of test results for the test system. For laboratories that the CAP accredits, we require scheduled calibration verification and verification of the analytical measurement range, which can be fulfilled by the laboratory's participation in the CAP Calibration Verification and Linearity (CVL) Program. The requirement for ongoing calibration verification provides an excellent example of how CLIA requirements help to provide post-approval performance monitoring for laboratories, FDA, and device manufacturers.

c. What timeframe do you consider when evaluating "real-world clinical use" performance?

CAP Response: *Under CLIA, laboratories must perform quality control at least daily when an assay is used and ongoing calibration verification of assays every 6 months. In addition, at least twice yearly, CAP accredited laboratories must participate in external proficiency testing or conduct alternative performance assessments to demonstrate the continued accuracy of test results. This timeline is also appropriate for AI-enabled medical devices.*

2. **Real-World Evaluation Methods and Infrastructure**
    a. What tools, methodologies, or processes are you currently using to proactively monitor AI-enabled medical device performance post-

deployment?

CAP Response: *Deep learning models for pathology images often require supplemental explainability tools, such as heatmaps, to enable expert verification, whereas classical statistical models are inherently interpretable. Some AI models are directly verifiable by a pathologist. For example, for an AI model used for cancer detection on histopathologic images, the pathologist can directly assess whether the AI findings align with established morphologic features. However, there are some AI models in which verification may be technically feasible but costly, requiring additional laboratory testing. For example, an AI system may indicate that prostate cancer has a high likelihood of harboring a specific mutation. The pathologist cannot confirm or refute this prediction by histopathologic examination alone. Verification by molecular testing is required.*

3. **Post-market Data Sources and Quality Management**
   a. What data sources do you typically use for ongoing performance evaluation (e.g., electronic health records, device logs, patient-reported outcomes)?

CAP Response: *Standardized performance monitoring of commercial testing has been largely dependent on uniform challenge tests applied globally, e.g., identical specimens shipped to multiple testing locations with comparison of results. However, AI differs from most clinical testing in its dependence on the statistical details of local data used as inputs. This means that the performance of local AI against external challenge tests may not be indicative of local performance, and that AI is susceptible to local data drift that may not be evident in external challenge tests. Therefore, best practices for AI performance monitoring should be standardized in their form and application but should evaluate both the performance of AI in the local context and the stability of local input data.*

As noted in response to question 1a, under CLIA, laboratories implement measures to detect drift and decreased performance in assays through daily quality control (QC) testing, calibration verification procedures, and monitoring positivity rates for qualitative outcomes looking for unexplained changes. Similar strategies are relevant for AI-based systems. For example, laboratories can trend data on local experience over time to monitor performance and ensure it does not change in ways that could harm patients. These practices—rooted in laboratory quality systems—provide a foundation for monitoring AI-enabled medical devices in real-world use.

5. **Human-AI Interaction and User Experience**
    a. How do clinical usage patterns and user interactions influence AI-enabled medical device performance over time based on your observations?

CAP Response: *User interactions encompass a wide range of behaviors and effects. In human-machine teaming scenarios, sometimes referred to as augmented intelligence,*[3] *machine augmentation tends to increase the proportion of agreement among human raters, such as pathologists. For example, a recent study introduced a method called CONTEST, which quantifies the effect of AI assistance on diagnostic agreement among pathologists.*[4] *Using data from prostate biopsy grading, the study demonstrated that AI augmentation significantly increased interobserver agreement, supporting the value of AI in enhancing diagnostic consistency. Similar results have been found in breast cancer*[5,6] *and neoplastic cellularity for lung cancer.*[7]

To enable local validation of these effects, vendors should provide detailed case tables from their internal validation studies to show how both human experts and AI systems performed on the same cases. Having access to detailed data from validation studies will enable institutions to compare and assess AI performance in their own clinical context.

    b. What design features, user training, or communication strategies have proven most effective for maintaining safe and effective use as systems evolve?

CAP Response: *The experience that pathologists and laboratory professionals implement through CLIA provides a helpful foundation for user training and communication strategies. The Clinical Laboratory and Standards Institute (a not-for-profit organization that develops consensus-based international laboratory standards and testing guidance) develops clear guidelines and best practices for maintaining testing standards. Laboratory accreditation and*

[3] Augmented Intelligence (AI) focuses on AI's assistive role, emphasizing that its design enhances human intelligence rather than replaces it. American Medical Association. October 21, 2025. Available at https://www.ama-assn.org/practice-management/digital-health/augmented-intelligence-medicine
[4] Olson BK, Rosenthal JH, Kappedal RD, Olson NH. CONTEST: *A generalization of ONEST to estimate sample size for predictive augmented intelligence method validation studies.* Journal of Pathology Informatics, 19:2025. https://doi.org/10.1016/j.jpi.2025.100519
[5] Jung M, Song SG, Cho SI, et al. *Augmented interpretation of HER2, ER, and PR in breast cancer by artificial intelligence analyzer: enhancing interobserver agreement through a reader study of 201 cases*. Breast Cancer Res. 2024 Feb 23;26(1):31. doi: 10.1186/s13058-024-01784-y.
[6] Mulder D, Shabaan A, Lacroix-Triki M, et al. *Use of artificial intelligence–assistance software for HER2-low and HER2-ultralow IHC interpretation training to improve diagnostic accuracy of pathologists and expand patients' eligibility for HER2-targeted treatment*. J Clin Oncol 43, 1014-1014(2025). DOI:10.1200/JCO.2025.43.16_suppl.1014
[7] Gertych A, Zurek N, Piaseczna N, et al.*Tumor cellularity assessment using artificial intelligence trained on immunohistochemistry-restained slides improves selection of lung adenocarcinoma samples for molecular testing.* American Journal of Pathology. 195:5, 907-922. https://doi.org/10.1016/j.ajpath.2025.01.009

proficiency testing also contribute to establishing and monitoring the quality of laboratory medicine; pathologists are directly involved in setting the standards and quality maintained by the CAP's accreditation, proficiency testing, and guidelines programs. These guidelines, best practices, and educational materials standardize laboratory testing and enable local laboratory generalist practitioners to appropriately monitor the quality of a wide variety of tests without being expert chemists, immunologists, or statisticians. Measuring and evaluating AI based upon this foundation of laboratory practice would help ensure reliance on the proven safety, quality, and accountability systems laboratories already use. The FDA's role in evaluating and clearing AI devices and ensuring transparency in device labels are vital elements to help pathologists operate within the CLIA model. The CAP also values ongoing collaborations and communications with FDA staff as part of the post-market monitoring environment.

* * * * *

To conclude, the CAP encourages the FDA to consider the unique insights from pathologists and laboratory medicine in its assessment of measuring and evaluating AI-enabled medical device performance in the real world. The CAP also notes that AI-enabled devices must be transparent and explainable for pathologists to best fulfill their duties concerning AI. The CAP stands willing to work with the FDA to provide our members' perspectives as CLIA-laboratory directors and section directors, and how this experience can be applied to the measurement and evaluation of AI-enabled medical devices.

Thank you for the opportunity to submit these comments. The CAP looks forward to working with the FDA. Please direct questions on these comments to Han Tran at htran@cap.org.

College of American Pathologists
1001 G Street, NW, Suite 425W
Washington, DC 20001
202-354-7100