

December 1, 2025

Sara Brenner, M.D., MPH
Principal Deputy Commissioner
U.S. Food and Drug Administration
10903 New Hampshire Ave,
Silver Spring, MD 20993

Submitted Electronically

RE: FDA-2025-N-4203 Request for Public Comment: Measuring and Evaluating Artificial Intelligence-enabled Medical Device Performance in the Real-World

Dear Deputy Commissioner Brenner,

On behalf of our nearly 5,000 member hospitals, health systems and other health care organizations, our clinician partners — including more than 270,000 affiliated physicians, 2 million nurses and other caregivers — and the 43,000 health care leaders who belong to our professional membership groups, the American Hospital Association (AHA) appreciates the opportunity to provide comment on the Food and Drug Administration (FDA) request for public comment regarding the measurement and evaluation of artificial intelligence (AI)-enabled medical devices.

The AHA values the critical role that the FDA plays in ensuring the safety and effectiveness of medical devices, including AI-enabled medical devices. We also appreciate that the FDA is evaluating ways to update and improve review pathways, especially considering the rapid pace of technological advancement for AI-enabled tools. AI-enabled medical devices offer tremendous promise for improved patient outcomes and quality of life. At the same time, they also pose novel challenges — including model bias, hallucinations and model drift — that are not yet fully accounted for in existing medical device frameworks. AI tools are inherently designed to be agile and adaptive, taking in new data points, discerning patterns and continually updating to improve model accuracy. This is especially true for generative AI.

In general, the AHA supports AI policy frameworks that balance flexibility to drive market-based innovations with appropriate safeguards to protect privacy and patient safety. As the FDA considers future policy approaches to measuring and evaluating AI-enabled medical device performance, we encourage the agency to:



- Pursue risk-based post-deployment measurement and evaluation standards for AI-enabled medical device vendors and developers.
- Synchronize measurement and evaluation activities with existing frameworks.
- Align incentives and address infrastructure barriers to measurement and evaluation.

Below are our specific recommendations.

AI Growth and Use in Hospitals and Health Systems

There has been significant growth in applications and use cases for AI in health care, with the majority of the over 1,240 AI-enabled medical devices approved by the FDA receiving approval in just the last three years.¹ Similarly, hospitals and health systems have significantly expanded their use of AI applications and continue to identify new AI use cases.

While many AI applications used by hospitals and health systems are administrative in nature (like AI-based revenue cycle and scheduling tools), hospitals also are deploying AI-enabled medical devices to augment care delivery. AI-enabled device and software use is becoming especially prevalent in diagnostic imaging and radiology. By leveraging advanced algorithms and machine learning techniques, AI can quickly identify patterns and anomalies in medical images that might be missed by human examination. For instance, AI can detect subtle changes, otherwise undiscernible by human detection, in tissues through X-rays, MRIs and CT scans, which is crucial for early disease detection.

Pursue Risk-based Post-deployment Measurement and Evaluation Standards for AI-enabled Medical Device Vendors and Developers

Per section 201(h) of the Food, Drug and Cosmetic Act, certain AI applications are considered medical devices if they are “intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease.” For AI-enabled medical devices, the FDA has aligned clearance guidelines with the existing 510k, de novo and pre-market approval processes. Meaning that the clearance process for AI-enabled medical devices is dictated by the level of risk, intended use and indications, just as with other medical devices. The highest risk applications are often life-saving products and require the most stringent review and standard controls. While it makes sense to align the clearance of AI-enabled medical devices with this existing framework, there are some gaps for post-deployment evaluation.

The potential for bias, hallucinations and model drift demonstrates the need for measurement and evaluation after deployment. The FDA should update adverse

¹ <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices>

reporting mechanisms to provide a more nuanced approach for the unique factors that impact the model integrity of AI-enabled medical devices. While manufacturers are required to report adverse events through the Manufacturer and User Facility Device Experience tool as part of the Medical Device Reporting program, the existing reporting variables do not include details on AI-specific risks. For example, the reporting tool simply has categories for malfunctions, injury or death. The agency could consider adding reporting variables for algorithmic stability and when there is a shift between training data and the real-world population served. Conducting these data comparisons can help to identify AI drift over time.

We also encourage the FDA to pursue a risk-based approach to monitoring and evaluation activities, whereby factors for potential risk to quality and patient safety are accounted for in the measures and scope of monitoring. In the spirit of a risk-based approach, the FDA could consider adding requirements for manufacturers to conduct a range of monitoring, from periodic revalidation activities to ongoing surveillance, depending on the classification of the AI-enabled medical device. In developing such frameworks, the FDA should seek feedback from device makers, hospitals and other providers, as well as standards development groups.

At the same time, evaluation and monitoring activities should not be overly burdensome and resource-intensive. As the FDA considers approaches to measurement and evaluation, we encourage the agency to consider end-user burden and take steps to minimize it. A risk-based approach to measurement and evaluation could focus scarce resources, such as time, personnel and cost, on the highest-risk applications and also align with the approach of domestically accredited and international standards development groups.

Synchronize Measurement and Evaluation with Existing Frameworks

The FDA has established foundational frameworks that can be leveraged as the basis for future post-deployment evaluation standards. We encourage the FDA to leverage and synchronize with these frameworks.

For example, measurement and evaluation should be aligned with the general total product lifecycle approach established by the FDA for medical devices. As such, post-market evaluation activities should be aligned with the existing clearance evaluation processes (de novo, 510k and pre-market approval). Developing separate frameworks could inadvertently add redundancy and inefficiency.

Synchronizing post-deployment measurement and evaluation activities also may support streamlining the clearance processes. Over 96% of AI-enabled medical devices are currently approved under the 510k process. However, the 510k process caps the number of indications for which applicants can seek approval at a given time. By its very nature, AI is adaptable, resulting in new capabilities that could augment the number of clinical indications for which a particular AI application is suited. Yet, the current 510k

clearance approach would generally require vendors and other developers to submit multiple clearance applications for the same technology. This approach may limit the ability to expand otherwise safe and effective uses of AI and result in added costs to vendors and providers alike, as application fees have often been passed down to end users through increased prices.

The AHA encourages the FDA to explore clearance pathways whereby vendors and other developers can provide detailed post-market evaluation and monitoring plans, including additional special controls, to streamline 510k clearances under fewer applications. We believe such an approach could shorten the time it takes to get evidence-based, safe uses of AI approved for use, benefiting patients and providers. It could also help reduce regulatory burden and cost.

The FDA also should clarify that measurement and evaluation standards would not apply to clinical decision support tools and administrative AI tools. The 21st Century Cures Act excluded certain AI tools from the purview of medical device oversight. For example, this statute excludes certain clinical decision support tools from the definition of “medical devices,” so long as they *support* clinical decision making, with the clinician independently reviewing recommendations. Therefore, clinical decision support tools should not be subject to evaluation and monitoring standards. Clinical decision support tools are often built to address a particular challenge for a specific period. Additional requirements can add barriers to delivering timely solutions and also may be impractical if the tool is only designed to support clinicians for a short timeframe (like six months). To maximize scarce resources, the FDA should focus monitoring and evaluation activities on higher-risk areas related to the diagnosis of disease or other conditions, or in the cure, mitigation, treatment or prevention of disease.

Align Incentives and Address Infrastructure Barriers to Measurement and Evaluation

While hospitals and health systems support evaluation and monitoring activities, vendors and developers must be responsible for the ongoing integrity of the tools they sell. Hospitals and health systems continually assess the strengths and limitations of AI models they use, but the “black box” nature of many AI systems can make it more challenging for hospitals and health systems to identify flaws in models that may affect the accuracy and validity of an AI tool’s analyses and recommendations. As such, post-market measurement and evaluation standards should be developed for vendors. Standards should include metrics for performance, thresholds for further evaluation, and communication requirements to end users related to ongoing performance.

Additionally, some hospitals (particularly rural, critical access and other safety net hospitals) may not have the staff or resources to support governance structures and ongoing measurement activities. As such, many of these health care providers have not deployed AI tools. Resource and infrastructure barriers can exacerbate the “digital divide” in certain geographies, where rural and other underserved areas have less

The Honorable Sara Brenner

December 1, 2025

Page 5 of 5

access to digital services, including AI tools for clinicians and patients. While addressing incentives and broader infrastructure barriers to AI is not the sole responsibility of the FDA, we encourage cross-agency collaboration to develop training, technical assistance and potential grant funding opportunities to support educational efforts and governance activities for AI health tools. This includes education and resources to support evaluation and monitoring activities. Cross-agency coordination could include collaboration among Health and Human Services, the FDA, the Federal Communications Commission, the Department of Commerce, the Department of Agriculture and the Department of Education.

We look forward to working with the FDA on frameworks to support the measurement and evaluation of AI-enabled medical device performance in the real world. Please contact me if you have questions, or feel free to have a member of your team contact Jennifer Holloman, AHA director of health IT policy, at jholloman@aha.org.

Sincerely,

/s/

Ashley Thompson
Senior Vice President
Public Policy Analysis and Development