Title

The CODEX action incubator: a consensus-driven approach to identify and implement diagnostic excellence measures in the context of artificial intelligence

Journal

Diagnosis, 0(0)

ISSN

2194-8011

Authors

Rosner, Benjamin
Hammer, Molly
Tabacco, Aaron
et al.

Publication Date

2026-08-13

DOI

10.1515/dx-2026-0112

Opinion Paper

Benjamin Rosner*, Molly Hammer, Aaron Tabacco, Julia Adler-Milstein and Sumant R. Ranji

The CODEX action incubator: a consensus-driven approach to identify and implement diagnostic excellence measures in the context of artificial intelligence

https://doi.org/10.1515/dx-2026-0112
Received June 15, 2026; accepted July 18, 2026;
published online August 13, 2026

Abstract: Diagnostic errors are a substantial source of patient harm. As artificial intelligence (AI) integrates into clinical workflows, opportunities are emerging to assess their impacts on diagnostic excellence (DxEx). The Coordinating Center for Diagnostic Excellence (CODEX) at the University of California San Francisco established the Action Incubator to translate research advances in DxEx into tangible strategies for improving diagnosis. The September 2025 in-person inaugural Action Incubator convened 30 multidisciplinary stakeholders representing health systems, patient advocacy, industry, and policy groups. Through structured discussions and breakout sessions, participants identified AI scribes as a near-term, scalable use case for evaluating AI's impact on diagnosis not only because of their widespread adoption, but – as supported by cognitive load theory – because of their potential to reduce cognitive burden and allow clinicians to focus more on diagnosis. A modified Delphi process was used to prioritize candidate measures based on feasibility, and impact. Participants generated 17 candidate measures of AI scribe impact on DxEx. Consensus was reached on two priority metrics as functions of AI scribe usage rates by primary care physicians: (1) Timely follow-up of abnormal test results related to breast and colorectal cancer screening and (2) Patient-reported diagnostic experience. Participating health systems will pilot these measures using electronic health record audit logs and patient surveys. The first CODEX Action Incubator developed a pragmatic, consensus-driven framework for measuring the impact of AI on DxEx. Future annual Action Incubators will take up timely, actionable topics related to DxEx.

Keywords: diagnostic excellence; artificial intelligence; quality measurement; patient safety; healthcare quality indicators

***Corresponding author: Benjamin Rosner**, MD, PhD, Coordinating Center for Diagnostic Excellence, Division of Clinical Informatics and Digital Transformation, University of California San Francisco, 10 Koret Way, Room K-301, UCSF DOM CLIIR Box 0737, San Francisco, CA 94143, USA; and Division of Hospital Medicine, University of California San Francisco, San Francisco, CA, USA, E-mail: Benjamin.rosner@ucsf.edu. https://orcid.org/0000-0003-3609-6481
Molly Hammer, Coordinating Center for Diagnostic Excellence, Division of Clinical Informatics and Digital Transformation, University of California San Francisco, San Francisco, CA, USA. https://orcid.org/0009-0000-2097-2585
Aaron Tabacco, Department of Medicine, University of California San Francisco, San Francisco, CA, USA. https://orcid.org/0009-0002-9479-8203
Julia Adler-Milstein, Division of Clinical Informatics and Digital Transformation, University of California San Francisco, San Francisco, CA, USA. https://orcid.org/0000-0002-0262-6491
Sumant R. Ranji, Coordinating Center for Diagnostic Excellence, Division of Clinical Informatics and Digital Transformation, University of California San Francisco, San Francisco, CA, USA; and Division of Hospital Medicine, University of California San Francisco at San Francisco General Hospital, San Francisco, CA, USA. https://orcid.org/0000-0002-1941-0005

Introduction

Diagnostic errors affect an estimated 12 million adults annually in the United States ambulatory care setting alone, with approximately half of these errors carrying the potential for harm [1]. The Coordinating Center for Diagnostic Excellence (CODEX) at the University of California San Francisco (UCSF) was established in 2024 with a mission of facilitating activities that result in measurable improvement in diagnostic quality, safety, and equity. CODEX considers diagnostic excellence (DxEx) to be the state achieved when clinicians and healthcare systems partner with patients to make accurate, timely, and equitable diagnoses, while using resources efficiently and managing uncertainty. Despite growing recognition of diagnostic safety as a priority, measuring and improving DxEx remains a persistent challenge for healthcare systems. However, the rapid integration

of artificial intelligence (AI) into clinical practice promises to help address this challenge. AI-based applications such as ambient scribes, predictive algorithms, and clinical decision support tools are being deployed across healthcare systems at unprecedented scale. While technologies such as AI scribes have been shown to reduce clinician task load and cognitive burden [2–4], whether the use of these tools creates cognitive bandwidth that results in improved diagnosis remains to be established [5]. A key priority for CODEX is to address a fundamental question: How can we ensure that AI helps to measurably improve diagnostic excellence?

This paper describes the concept and development of the CODEX Action Incubator, the mechanism by which CODEX seeks to engage interdisciplinary stakeholders and drive improvements in DxEx. We further describe the inaugural gathering and the resulting metrics of DxEx being put into action by participating health systems to determine whether DxEx benefits are being realized from general purpose AI tools used in clinical care.

CODEX Action Incubator: concept and development

The CODEX Action Incubator was modeled after other successful interdisciplinary convening approaches, including the Aspen Institute's Ideas to Action Incubator [6] and the Rural Action Roundtable on Equity [7]. The CODEX Action Incubator seeks to bring together a small, diverse group of stakeholders with the directive of moving ideas to action. Unlike traditional research consortia, the Action Incubator was designed to emphasize actionability by identifying actions that could be taken to translate scientific studies pertaining to diagnostic excellence into tangible next steps that drive enhancements in healthcare. What differentiates an Action Incubator from other approaches to improving DxEx (e.g., research or education) is its call for healthcare organizations to commit to taking measurable steps toward DxEx in a specific area.

The Action Incubator emphasizes five key principles: actionability, measurability, applicability (actions and findings should be broadly applicable across a breadth of healthcare settings), longitudinality (activity should support ongoing improvement), and sustainability. To identify key goals for recurring Action Incubators that would resonate with the DxEx and broader quality and patient safety communities, we began with a landscape analysis that included semi-structured interviews with 14 national thought leaders in the areas of diagnosis, policy, patient safety, and research. The interviews explored defining success criteria for the Action Incubators, suggested composition of Action Incubator participants, specific outcomes and products, approaches for sustainable engagement, and potential topics for the inaugural gathering.

A common denominator from these stakeholder interviews was that a smaller in-person gathering would be better for productive output and for participant engagement. Initiating an Action Incubator with an in-person meeting with a limited set of stakeholders was felt to be optimal, but participants would need to represent a broad diversity of the DxEx stakeholder and DxEx-adjacent communities. Furthermore, a successful Action Incubator would not only lead to DxEx initiatives with measurable outcomes in health systems but would catalyze sustained stakeholder engagement. Uniformly, stakeholder interviewees felt that the opportunities for generative AI in DxEx represented the most important, timely topic for the inaugural Action Incubator. Interviewees also recognized, however, that: 1. AI scribes could have unintended consequences many of which, at the time, could not yet be envisioned [8], and 2. Not every health system would have AI tools deployed yet. Therefore, their recommendation was not to expect or request new AI deployments by participating health systems as a condition of participation, but rather to examine DxEx in the context of tools already deployed, with the most promising candidate being the AI scribe.

Inaugural CODEX Action Incubator

Topic

Although AI scribe tools are primarily meant for documentation, interviewees posited that they could have secondary beneficial impacts on DxEx by decreasing the baseline level of clinician cognitive burden and therefore allowing clinicians more cognitive bandwidth for other activities including diagnosis. The premise for this hypothesis – that AI scribes decrease cognitive burden – has since been borne out in the literature [2, 3], and this mechanism is also supported by cognitive load theory, describing that within the constraints of a clinician's finite working memory capacity distributed between extrinsic and internal loads, when extrinsic load is decreased, the clinician may redistribute that capacity to other clinical tasks [5] (Figure 1).

Participants and initial convening

CODEX kicked off its inaugural Action Incubator with an in-person, 1.5 day meeting on September 11–12, 2025 at UCSF. Thirty national healthcare leaders participated, including operational and research representatives from health systems throughout the U.S. (academic medical centers, community health systems, and safety net systems), patient

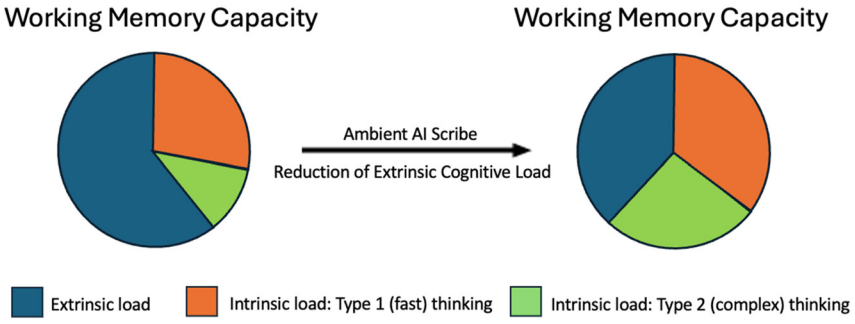

Figure 1: Conceptual model of the role of the AI scribe in diagnostic excellence. As extrinsic load on a clinician decreases (such as through the use of an ambient AI scribe), a greater proportion of the clinician's finite working memory capacity may be allocated to Type 1 (fast) and Type 2 (analytical) thinking, including following up on abnormal test results (adapted from AHRQ [5]).

Table 1: Action Incubator candidate DxEx metrics relating to the use of AI scribes.

Item	Required data source(s)
Time to follow up on abnormal test results	EHR audit logs, lab information system
AHRQ SOPS diagnostic safety supplemental item set for the SOPS Medical Office Survey [9]	Provider survey
AHRQ Quality Indicator, missed red flag SaferDx e-measure 1 (cancer, chest imaging) [10]	EHR/PACS
AHRQ Quality Indicator, missed red flag SaferDx e-measure 2 (colorectal cancer) [10]	EHR, lab information system
Closing loop on referrals	EHR (May not be available if referrals are outside of a health system)
Proportion of patients with late-stage colorectal cancer diagnosis	EHR
Proportion of patients with late-stage lung cancer diagnosis	EHR
Patient perspectives on diagnostic communication and timeliness	Patient survey
Diagnostic accuracy	EHR (although EHR may not consistently have a final diagnosis)
Diagnostic disposition accuracy	EHR
Diagnostic guideline adherence	EHR or registry
Proposed supplemental question to OAS CAHPS [11]: patient trust in diagnosis	Proposed new question embedded in OAS CAHPS survey
Clinician diagnostic uncertainty	New field in EHR or in text of clinician documentation
Patient as partner: patient asked about use of AI in self-assessment prior to encounter	Scribe
Time from symptom onset to pathology-confirmed diagnosis	Patient, EHR
Time from first medical contact to diagnosis	Patient survey, EHR
Time from symptom onset to first contact with healthcare system	Patient, EHR

AHRQ, Agency for Healthcare Research and Quality; DxEx, diagnostic excellence; EHR, electronic health record; OAS CAHPS, Outpatient Ambulatory Surgery Consumer Assessment of Healthcare Providers and Systems; PACS, picture archiving and communication system.

advocacy organizations, medical malpractice experts, payers and insurers, industry and vendors, and the federal government. The goal of the kickoff meeting was to bring together interdisciplinary stakeholders to drive measurable improvements in diagnosis and answer the question: How do we measure whether *current* use of ambient scribes is improving DxEx? This could serve as a model for ongoing work assessing the impact of AI on DxEx outcomes and catalyze efforts to enhance this impact.

The meeting was organized into two parts, each with a keynote presentation to provide context. *Part 1 –Measuring Diagnostic Excellence in the Current State of Widely Adopted Ambulatory Care AI Scribes* convened with a keynote on learnings from the use of AI in the field of radiology, and *Part 2 – Framing for the Future: Measuring Diagnostic Excellence when AI is Widely Embedded in the electronic health record (EHR)* opened with a keynote from UCSF's Chief AI Officer. Each part consisted of professionally moderated discussion and breakout group sessions to ideate on candidate metrics.

Part 1 takeaways included: 1. AI can supplement clinicians' work (e.g., speeding the rate of image interpretation among radiologists), but may not, without the clinician, have the broader patient context that is often necessary for medical decision-making, and 2. While it is important to evaluate AI performance on diagnostic accuracy, it is equally important to evaluate how AI fits into the clinical workflow.

Breakout groups considered measurement opportunities across several diagnostic excellence frameworks: diagnostic accuracy, diagnostic guideline adherence, diagnostic disposition accuracy (e.g., is a patient correctly discharged from the emergency department home rather than

being admitted?), and across the crosscutting concept of diagnostic timeliness. These breakout groups generated a list of candidate metrics for DxEx that could be improved by AI scribe use (see Table 1). In so doing, contributors were asked to consider measurement feasibility, scale of potential patient impact, causal association of the outcome with AI, and implementation effort.

Part 2 takeaways included: 1. Once widely deployed into the EHR, AI tools should be evaluated against current healthcare delivery rather than idealized standards, and 2. Metrics for DxEx would likely differ by care setting including inpatient, ambulatory, and emergency department, but that unlocking the full potential of AI for decision support and autonomous care models would require strong governance in all settings.

Consensus process for identifying candidate measures

Following the in-person meeting, CODEX aggregated and reviewed the proposed measures from group discussions (Table 1). The initial set of metrics included 17 proposed measures for the current state (AI scribes), and six additional measures for the future state (the subject of a separate future manuscript) when AI is fully embedded in the EHR.

Next, we used a modified Delphi approach to achieve consensus on priority metrics through iterative asynchronous surveys alternating with three virtual meetings. Early in the process, the group prioritized follow up on abnormal test results as a high-priority need in DxEx that would have measurement feasibility, large-scale impact, causal association with the use of AI scribes (mechanism in Figure 1), and implementation efficiency. Similarly, the group felt that it was important to capture the patient experience and adapted drafted survey questions under development by Patients for Patient Safety, one of the organizations represented at the Action Incubator. At the conclusion of the modified Delphi process, the group achieved consensus on the metrics shown in Table 2. One health system, a children's hospital, for which two of the metrics would not be

Table 2: CODEX Action Incubator current state consensus metrics.

Metric name	Metric description	Measurement source
Clinician timely follow up on abnormal test results		
Rate of timely follow-up on abnormal screening mammograms for breast cancer detection [12][a]	AHRQ Quality Indicator 5: an electronic Clinical Quality Measure (eCQM) that reports the percentage of female patients aged 40–75 years with at least one abnormal screening (BI-RADS 0) or screening-to-diagnostic (BI-RADS 4, 5) mammogram during the measurement period (i.e., calendar year) who received timely diagnostic resolution defined as either follow-up imaging with negative/benign/*probably benign* results or a breast biopsy or diagnostic excision procedure within 60 days after their index (i.e., first) abnormal screening mammogram.	EHR audit logs to assess for a follow up order being placed for diagnostic imaging or for breast biopsy/diagnostic excision procedure within 60 days after index abnormal screening mammogram. Examine EHR logs for flags indicating the degree to which the screening physician uses AI scribes.
Rate of timely follow-up on positive stool-based tests for colorectal cancer detection [13][a]	AHRQ Quality Indicator 4: an electronic Clinical Quality Measure (eCQM) that reports the percentage of patients aged 45–75 years with at least one positive stool-based colorectal cancer screening test (i.e., high-sensitivity guaiac fecal occult blood test, fecal immunochemical test, or Cologuard) during the measurement period (i.e., calendar year) who completed a colonoscopy within 180 days after their index (i.e., first) positive stool-based test result date.	EHR audit logs to assess for a follow up order being placed for colonoscopy within 180 days after index positive stool-based screening test result date.[b] Examine EHR logs for flags indicating the degree to which the screening physician uses AI scribes.
Patient perspective		
Patient survey: diagnostic communication	After your recent primary care visit, were the results of your [your child's] diagnostic tests communicated to you in a way you could easily understand?	Collected from survey responses. Examine EHR logs for flag indicating whether AI scribe was used in relevant patient encounter.
Patient survey: diagnostic timeliness	After your recent primary care visit, do you feel that you [your child] received a timely diagnosis?	Collected from survey responses. Examine EHR logs for flag indicating whether AI scribe was used in relevant patient encounter.

eCQM, electronic clinical quality measure; EHR, electronic health record. [a]One Action Incubator health system participant is a children's hospital for which these measures do not apply based on age criteria. The children's hospital intends to implement AI, to assess when a parent with a child for whom there are unfulfilled lab tests is on campus, and to push a nudge to complete pending lab tests while a parent is on campus. [b]Although the AHRQ Quality Indicator is a measure of colonoscopy being *completed* within 180 days, we will examine whether an order was *placed* within 180 days as an assessment of actions entirely within the control of the clinical provider.

applicable, developed a unique metric for its own implementation.

Next steps: piloting the framework across participating health systems

As a next step, Action Incubator health system participants will pilot the selected measures at their institutions. For the first two metrics, they will examine through the EHR audit logs whether the expected follow up action on abnormal test results was taken in the appropriate time frames. For metrics on patient perspective, participating health systems will deploy patient survey questions about diagnostic communication and timeliness. The goal for all metrics is to understand whether use of AI scribes is associated with changes in rates of follow-up on abnormal test results and patient perceptions of diagnostic communication and timeliness. If AI scribe use is associated with improved follow up and patient experience, it will support the hypothesis that AI scribes help optimize clinician cognitive load, and health systems will have data to support promoting adoption of these AI tools among clinicians (especially those whose scribe utilization and follow up on the selected screening tests is low). If no association is found, then the next actionable steps may be for health systems to examine other methods of improving abnormal test result follow up – including AI and non-AI tools – and consider investing in AI tools for direct diagnostic support.

Conclusions

The inaugural CODEX Action Incubator revealed several important insights for measuring DxEx in the context of AI implementation.

First, AI offers opportunities as both an intervention (e.g., scribe) and as a tool for measurement (e.g., chart reviewer). At the time of the Action Incubator, AI scribes mostly lacked integrated clinical decision support, so a conceptual model based on cognitive load reduction and its impact on DxEx had to be considered. However, in a future state, when clinical decision support is embedded with AI in the EHR, health systems may continue to focus on follow up on abnormal test results (or other measures) but be better positioned to assess the impact of AI on DxEx more directly. As a tool for measurement, AI could transform manual chart review and adjudication, commonly used in DxEx, by making it more scalable than possible with humans alone. Because the field of AI is moving so rapidly, new opportunities for AI in DxEx, not yet envisioned at the Action Incubator, are sure to emerge.

Next, there is a healthy dynamic between feasibility and impact in the actions that can be taken following an Action Incubator. By design, the measurement of DxEx among Action Incubator participants was intended to balance feasibility and implementation effort, while simultaneously having as large an impact as possible on the populations of these systems. Therefore, while some candidate metrics may have been highly feasible, they were excluded in the modified Delphi process due to the low volume of patients they might impact.

Finally, all participants acknowledged that as with many assessments, there are attribution challenges. Health systems are complex entities with many moving parts. While we hope to assess for association between AI scribe use and DxEx, this is intended to be a pragmatic assessment of the impact of AI scribes on DxEx and is not designed to establish causation. Nevertheless, if carried out correctly, this may be a substantial step forward in the way that we measure improvements in DxEx at scale from AI adoption.

The CODEX Action Incubator represents a new approach to developing consensus-based metrics for an emerging challenge in healthcare quality and provides a framework for health systems to measure diagnostic excellence as AI tools are deployed. By bringing together diverse stakeholders, the initiative identified practical measures that balance feasibility with impact.

The consensus on selecting Agency for Healthcare Research and Quality (AHRQ) Quality Indicators for abnormal test result follow-up by the Action Incubator participants provides health systems with readily implementable metrics that offer clear opportunities for AI-supported improvement. The patient voice dimensions will incorporate perspectives often overlooked in diagnostic quality measurement.

This work represents a mid-point in the first of UCSF CODEX's Action Incubators. Once completed, we intend to report the results of the measures implemented at participating health systems, not only as a model for other health systems, but also as a decision point for health system leaders to either promote AI scribe use if there is a positive association with these outcomes – particularly among low utilizers – or to plan for dedicated tools for AI diagnostic decision support. Moving forward, CODEX's goal is to hold Action Incubator meetings approximately annually, inviting national thought leaders to address the most pressing and timely issues to the field of DxEx. CODEX welcomes public input, including topic ideas for subsequent Action Incubators. These can be sent to CODEX@ucsf.edu.

Acknowledgments: The authors thank the following Action Incubator participants for their contributions: Lauren Alderson MD, OCHIN; Andrew Auerbach MD, MPH, UCSF; Karen Cosby MD; Jason Cunningham DO, West County

Health Centers; Andrea Downing, The Light Collective; Rob El-Kareh MD, MS, MPH, UC San Diego; Ellen Flynn RN, MBA, JD, Vizient PSO; Mark Friedberg MD, MPP, Blue Cross Blue Shield of Massachusetts; Ethan Goh MD, MS, Stanford University; Veronique Grenon, The Doctors Company; David Larson MD, MBA, Stanford University; Zach Lipton PhD, Abridge; Vincent Liu MD, MSc, Kaiser Permanente; Carl Macrae PhD, University of Nottingham; Kristin Miller DrPH, MSL, MSPH, Medstar; Rebecca Mishuris MD, MS, MPH, Mass General Brigham; Sara Murray MD, MAS, UCSF; Andrew Olson, MD, University of Minnesota; Christopher Ott, MD, HCA Healthcare Physician Services; Michelle Schreiber MD, Centers for Medicare and Medicaid Services; Anjana Sharma MD, MAS, UCSF; Sue Sheridan MIM, MBS, DHL, Patients for Patient Safety US; Craig Umscheid MD, MS, MedStar; Kevin Wang MD, Suki; Kevin Watson Jr. MD, Akron Children's Hospital; TJ Wolski DO, Akron Children's Hospital; and Daniel Yang MD, Kaiser Permanente. We also acknowledge UCSF staff Stephanie Chuc MHA, Erin Hartman MS, Caroline Jens MPH, and Mika Rivera MPS. Project PIVOT provided essential input on patient experience survey questions. Special acknowledgment is given to the patient advocates who ensured that the patient voice remained central to discussions.

Research ethics: IRB not applicable.

Informed consent: Not applicable.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Use of Large Language Models, AI and Machine Learning Tools: The ideas and organization of the abstract were ours, but we used ChatGPT to improve conciseness of the abstract. The authors subsequently reviewed and edited the abstract and take full responsibility for it.

Conflict of interest: BR is an employee of both UCSF and Ambience Healthcare. SR reports funding from AHRQ for diagnostic error research and reports royalties from UpToDate for authorship of an online textbook chapter.

Research funding: CODEX is supported by a grant from the Gordon and Betty Moore Foundation. There was no specific funding in support of this manuscript or the ongoing Action Incubator measurement activities.

Data availability: Not applicable.

References

1. Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014;23:727–31.
2. Pelletier JH, Watson K, Michel J, McGregor R, Rush SZ. Effect of a generative artificial intelligence digital scribe on pediatric provider documentation time, cognitive burden, and burnout. JAMIA Open 2025; 8:ooaf068.
3. Olson KD, Meeker D, Troup M, Barker TD, Nguyen VH, Manders JB, et al. Use of ambient AI scribes to reduce administrative burden and professional burnout. JAMA Netw Open 2025;8:e2534976.
4. Lukac PJ, Turner W, Vangala S, Chin AT, Khalili J, Shih YT, et al. Ambient AI scribes in clinical practice: a randomized trial. NEJM AI 2025;2. https://doi.org/10.1056/aioa2501000.
5. Knees M, Raffel KE, Kissler M, Burden M, Porter S, Schnipper J, et al. Cognitive load theory and its impact on diagnostic accuracy; 2024. Available at: http://file:///Users/brosner/Downloads/dxsafety-cognitive-load-theory%20(2).pdf.
6. Aspen Institute. Hatching new ideas to deliver health care to the last mile – Aspen Institute. Ideas to action incubator; 2015. Available at: https://www.aspeninstitute.org/blog-posts/hatching-new-ideas-deliver-health-care-last-mile/.
7. Aspen Institute. Rural action roundtable on equity – Aspen CSG. Rural action roundtable on equity; 2022. Available at: https://www.aspencsg.org/rural-action-roundtable-on-equity/.
8. Nong P, Neprash HT. Unintended consequences of using ambient artificial intelligence scribes for billing. JAMA Health Forum 2026;7: e255771.
9. Agency for Healthcare Research and Quality. Diagnostic safety supplemental items for medical office SOPS | Agency for Healthcare Research and Quality. Diagnostic safety supplemental items for medical office SOPS; 2021. Available at: https://www.ahrq.gov/sops/surveys/medical-office/supplemental-items/diagnostic-safety.html.
10. Agency for Healthcare Research and Quality. AHRQ Quality Indicators; 2024. Available at: https://qualityindicators.ahrq.gov/Downloads/Resources/Diagnostic%20Excellence%20Measure%20for%20RFI.pdf.
11. Outpatient and Ambulatory Surgery CAHPS Survey. Outpatient and ambulatory surgery CAHPS survey. Available at: https://oascahps.org/.
12. AHRQ. AHRQ Quality Indicator: rate of timely follow-up on abnormal screening mammograms for breast cancer detection. Partnership for quality measurement; 2025. Available at: https://p4qm.org/measures/4700e.
13. Agency for Healthcare Research and Quality. AHRQ Quality Indicator: rate of timely follow-up on positive stool-based tests for colorectal cancer detection. Partnership for quality measurement; 2025. Available at: https://p4qm.org/measures/4705e.