2025 Pre-Rulemaking Measure Review
Preliminary Assessment

MUC ID	Title
MUC2025-042	Rate of Timely Follow-up on Abnormal Screening Mammograms for Breast Cancer Detection
Measure Steward & Developer	**Proposed CMS Programs**
Brigham and Women's Hospital	Merit-based Incentive Payment System (MIPS) Link: Merit-based Incentive Payment System (MIPS)

Measure Overview

Rationale: Breast cancer is the second most common cause of cancer deaths among women in the United States. In 2025, around 42,170 women will die from breast cancer, and an estimated 316,950 new cases of invasive breast cancer will be diagnosed.[1]

Breast cancer survival is dependent upon cancer stage at diagnosis. Approximately 99% of women diagnosed with early stage breast cancer live for 5 years or more.[2] However, this applies to only about 32% of those diagnosed at the most advanced stage.

Noninvasive mammographic screening is the primary screening modality used to detect breast cancer. Delays in diagnostic follow-up after abnormal mammographic screening results increase the risk of diagnosing cancer at a more advanced stage.[3]

National screening guidelines recommend that women with abnormal screening mammogram results (BI-RADS 0, 4, or 5) undergo additional follow-up imaging via diagnostic mammography, magnetic resonance imaging (MRI), and/or ultrasound.[4,5,6,7] While it is recommended that patients with a benign follow-up imaging result return to routine screening, those with abnormal results (BI-RADS 4 or 5) should have diagnostic samples extracted (e.g., via core needle biopsy, fine needle aspiration, or surgical excision) from a suspicious area to evaluate for cancer.[4]

Expert-based quality measure programs support the need to establish a reasonable timeframe that encompasses this multi-step process. According to the Centers for Disease Control and Prevention

[1] Key Statistics for Breast Cancer. American Cancer Society. Updated May 5, 2025. Accessed September 12, 2025. https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html.
[2] Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on 2021 submission data (1999–2020): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Updated June 2024. Accessed July 2024. www.cdc.gov/cancer/dataviz.
[3] McCarthy AM, Kim JJ, Beaber EF, et al. Follow-Up of Abnormal Breast and Colorectal Cancer Screening by Race/Ethnicity. Am J Prev Med. 2016; 51(4):507-512. doi:10.1016/j.amepre.2016.03.017. PMID: 27132628.
[4] Sickles E, D'Orsi CJ. ACR BI-RADS follow-up and outcome monitoring. In: D'Orsi CJ, ed. ACR BI-RADS atlas, breast imaging reporting and data system. Reston, VA: American College of Radiology Reston; 2013:5-67. https://www.acr.org/-/media/ACR/Files/RADS/BI-RADS/BIRADSFAQ.pdf.
[5] Monticciolo DL, Malak SF, Friedewald SM, et al. Breast Cancer Screening Recommendations Inclusive of All Women at Average Risk: Update from the ACR and Society of Breast Imaging. J Am Coll Radiol. 2021; 18(9):1280-1288. doi:10.1016/j.jacr.2021.04.021. PMID: 34154984.
[6] US Preventive Services Task Force, Nicholson WK, Silverstein M, et al. Screening for Breast Cancer: US Preventive Services Task Force Recommendation Statement [published correction appears in JAMA. 2024 Sep 30. doi: 10.1001/jama.2024.19851]. JAMA. 2024; 331(22):1918-1930. doi:10.1001/jama.2024.5534. PMID: 38687503.
[7] Esserman LJ, Joe BN, et al. Diagnostic evaluation of suspected breast cancer. UpToDate. Updated October 31, 2023. Accessed October 31, 2024. https://www.uptodate.com/contents/diagnostic-evaluation-of-suspected-breast-cancer?search=birads&source=search_result&selectedTitle=2%7E13&usage_type=default&display_rank=2#H24.

Version 1.0 | December 2025 | *The analyses upon which this publication (or document) is based were performed under Contract Number 75FCMC23C0010, entitled, "National Consensus Development and Strategic Planning for Health Care Quality Measurement," sponsored by the Department of Health and Human Services, Centers for Medicare & Medicaid Services. Restricted:* Use, duplication, or disclosure is subject to the restrictions as stated in Contract Number 75FCMC23C0010 between the Government and Battelle.

Measure Overview
(CDC) National Breast and Cervical Cancer Early Detection Program (NBCCEDP), breast cancer screening to diagnostic resolution should occur within 60 days.[8] It is also expected that over 90% of women complete diagnostic resolution after an abnormal screening mammogram.[8,9] Published literature shows that long wait times to diagnostic evaluation are associated with increased tumor size and lymph node metastases in patients with delays exceeding 12 weeks.[10,11,12] In particular, invasive triple negative breast cancers have been shown to double in size in <60 days.[13] Differences in diagnostic follow-up rates after abnormal screening mammograms are reported in the literature. A 2021 systematic review reported rates of failure to follow-up on abnormal screening mammograms ranging from 7.2-33%.[14] A 2024 study on the American College of Radiology's National Mammography Database (NMD) observed that only 66.4% of 2.9 million abnormal screening mammograms (BI-RADS 0) documented from 2008-2021 had diagnostic follow-up. In this cohort, women with no family history of breast cancer had lower follow-up rates, and Black and Native American women had lower overall follow-up rates and lower biopsy rates. Rural and community hospital-affiliated facilities had longer median times to biopsy.[15] The variability in follow-up rates in the NMD and existing literature imply the existence of barriers limiting mammography facilities from carrying out complete diagnostic resolution within a timely manner for all patients. This electronic clinical quality measure (eCQM) can be used to address quality assessment gaps by monitoring timeliness and completeness of care in medical facilities looking to improve the breast cancer screening and diagnostic process.
CMS-provided program rationale: CMS is considering adding this measure to the MIPS quality measure set as a new measure for future performance years. MIPS does not have any related measures that examine timely follow-up for abnormal screening mammograms; therefore, the quality of patient care benefits from the promotion of early detection of breast cancer through this measure. This measure is fully tested and developed at both the facility and clinician level. This process measure represents a gap in MIPS and CMS priority areas for diagnostic radiology, which has limited measures and digital measurement overall. Additionally, the measure may be considered for potential inclusion in the diagnostic radiology MIPS Value Pathway (MVP).
Description: This eCQM reports the percentage of female patients aged 40 to 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 within 60 days after their index (i.e., first) abnormal screening mammogram. Negative/benign/probably benign follow-up imaging was defined as diagnostic mammography, breast ultrasound or magnetic resonance

[8] DeGroff A, Royalty JE, Howe W, et al. When performance management works: a study of the National Breast and Cervical Cancer Early Detection Program. Cancer. 2014; 120 Suppl 16(Suppl 16):2566-2574. doi:10.1002/cncr.28817. PMID: 25099899.
[9] Miller JW, Hanson V, Johnson GD, Royalty JE, Richardson LC. From cancer screening to treatment: service delivery and referral in the National Breast and Cervical Cancer Early Detection Program. Cancer. 2014; 120 Suppl 16(0 16):2549-2556. doi:10.1002/cncr.28823. PMID: 25099897.
[10] Olivotto IA, Gomi A, Bancej C, et al. Influence of delay to diagnosis on prognostic indicators of screen-detected breast carcinoma. Cancer. 2002; 94(8):2143-2150. doi:10.1002/cncr.10453. PMID: 12001110.
[11] Ganry O, Peng J, Dubreuil A. Influence of abnormal screens on delays and prognostic indicators of screen-detected breast carcinoma. J Med Screen. 2004; 11(1):28-31. doi:10.1177/096914130301100107. PMID: 15006111.
[12] Doubeni CA, Gabler NB, Wheeler CM, et al. Timely follow-up of positive cancer screening results: A systematic review and recommendations from the PROSPR Consortium. CA Cancer J Clin. 2018; 68(3):199-216. doi:10.3322/caac.21452. PMID: 29603147.
[13] Nakashima K, Uematsu T, Takahashi K, Nishimura S, Tadokoro Y, Hayashi T, Sugino T. Does breast cancer growth rate really depend on tumor subtype? Measurement of tumor doubling time using serial ultrasonography between diagnosis and surgery. Breast Cancer. 2019 Mar; 26(2):206-214. doi: 10.1007/s12282-018-0914-0. Epub 2018 Sep 26. PMID: 30259332.
[14] Reece JC, Neal EFG, Nguyen P, McIntosh JG, Emery JD. Delayed or failure to follow-up abnormal breast cancer screening mammograms in primary care: a systematic review. BMC Cancer. 2021; 21(1):373. Published 2021 Apr 7. doi:10.1186/s12885-021-08100-3. PMID: 33827476.
[15] Oluyemi ET, Grimm LJ, Goldman L, et al. Rate and Timeliness of Diagnostic Evaluation and Biopsy After Recall From Screening Mammography in the National Mammography Database. J Am Coll Radiol. 2024; 21(3):427-438. doi:10.1016/j.jacr.2023.09.002. PMID: 37722468.

Measure Overview
imaging (MRI) with BI-RADS ratings of 1, 2, or 3. Relevant diagnostic breast biopsy procedures were defined as core needle biopsy, fine needle aspiration, and surgical excision. Breast Imaging – Reporting and Data System (BI-RADS) ratings: 0-incomplete, 1-negative, 2-benign, 3-probably benign, 4-suspicious, 5-highly suggestive of malignancy.
Measure background: New measure; never reviewed by a Measure Applications Partnership (MAP) Workgroup or PRMR committee; never used in a Medicare program.
Numerator: Patients in the denominator population who received timely diagnostic resolution defined as negative/benign/probably benign follow-up imaging (BI-RADS 1, 2, 3) or breast biopsy within 60 days after the date of their index (i.e., first) abnormal screening (BI-RADS 0) or screening-to-diagnostic (BI-RADS 4, 5) mammogram. Extract the date of the first abnormal screening (BI-RADS 0) or screening-to-diagnostic (BI-RADS 4, 5) mammogram in the measurement period (i.e., calendar year) for each patient to define the index screening mammograms and index dates (i.e., start of the follow-up period) [value sets: “Screening Mammogram (Grouping)” OID 2.16.840.1.113762.1.4.1206.61; BIRADSCategories04And5 OID 2.16.840.1.113762.1.4.1206.67]. If documented, extract the first follow-up imaging (i.e., diagnostic mammogram, ultrasound, or MRI) with negative/benign/probably benign (BI-RADS 1, 2, 3) ratings within 60 days after the date of the index abnormal screening mammogram for each patient [value sets: “Diagnostic Mammography” OID 2.16.840.1.113762.1.4.1206.65; “Ultrasound of the Breast” OID 2.16.840.1.113883.3.3157.1902; “MRI of the Breast” OID 2.16.840.1.113883.3.3157.1903; BIRADSCategories12And3 OID 2.16.840.1.113762.1.4.1206.68]. If documented, extract the first breast biopsy procedure (i.e., core needle biopsy, fine needle aspiration, or surgical excision) within 60 days after the date of the index abnormal screening mammogram for each patient [value set: “Breast Cancer Biopsy and Surgical Excision” OID 2.16.840.1.113762.1.4.1206.66]. Patients that received negative/benign/probably benign follow-up imaging or breast biopsy within 60 days are included in the numerator population. **Exclusions:** N/A
Denominator: Female patients aged 40 to 75 years with an abnormal screening (BI-RADS 0) or screening-to-diagnostic (BI-RADS 4, 5) mammogram during the measurement period (i.e., calendar year). Only the first abnormal screening or screening-to-diagnostic mammogram (i.e., index screening test) is included in the measure calculation. Extract all abnormal screening mammograms (BI-RADS 0) and screening-to-diagnostic mammograms (BI-RADS 4, 5) during the measurement period (i.e., calendar year) [value sets: “Screening Mammogram (Grouping)” OID 2.16.840.1.113762.1.4.1206.61; BIRADSCategories04And5 OID 2.16.840.1.113762.1.4.1206.67]. Retain abnormal screening and screening-to-diagnostic mammograms where the patient was aged between 40 and 75 years on the date of the mammogram [value set "Birth Date" OID 2.16.840.1.113883.3.560.100.4]. Retain abnormal screening and screening-to-diagnostic mammograms where the patient was female [value set "ONC Administrative Sex" OID 2.16.840.1.113762.1.4.1]. Patients with at least one abnormal screening or screening-to-diagnostic mammogram are included in the denominator population. **Exclusions:** N/A **Exceptions:** N/A

Measure Overview	
Patients with at least one abnormal screening or screening-to-diagnostic mammogram are included in the denominator population. **Exclusions:** N/A **Exceptions:** N/A	
Substantive changes from prior version (if applicable): N/A	
Measure type: Process	**Measure is a composite:** No **Measure is digital and/or an eCQM:** Yes **Measure is a paired or group measure:** No
Level of analysis: Facility	**Data source(s):** Digital-Electronic Health Record (EHR) Data
Care setting(s): Hospital Outpatient Department (HOD), Ambulatory/office-based care	**Risk adjustment or stratification:** No
CBE endorsement status: Endorsed	**CBE endorsement history:** Endorsed with conditions during the Spring 2024 cycle. When the measure returns for maintenance (3 years), the measure developer should have: • Conducted additional validity testing (data element in additional EHR); and • Continued to monitor (e.g., qualitative assessments, empirical analyses) for unintended consequences (e.g., reduced access to mammography) during implementation.
Is measure currently used in CMS programs? No	**Measure addresses statutorily required area?** No

Evaluation

Meaningfulness

Importance	
Type of evidence:	Clinical Guidelines or U.S. Preventive Services Task Force (USPSTF) Guidelines; Empirical data; Peer-Reviewed Original Research; Peer-Reviewed Systematic Review [MUC Entry/Review Information Tool (MERIT) Submission Form]
Importance: As outlined in the literature cited for the measure rationale, early detection of breast cancer through routine mammographic screening has significantly reduced mortality and treatment costs. Studies show that most breast imaging facilities do not meet benchmarks for timely follow-up imaging and biopsy, but participation in quality measurement programs improves performance, especially in underperforming facilities. This eCQM has been developed to help facilities routinely assess and improve the timeliness of diagnostic resolution after abnormal mammograms, supporting better outcomes and more equitable care. During CBE endorsement review in 2024, the committee found the evidence supporting the importance of this measure to be sufficient.	
Rating: Met; Prior CBE Endorsement	

Conformance
Measure alignment with conceptual intent: This new measure is intended to calculate the rate of timely diagnostic resolution in facilities that perform mammographic screening after an abnormal screening mammogram to detect breast cancers. The measure numerator, denominator, and exclusions for the measure scores are defined and support the intent of the measure. The measure aligns with MIPS objectives to 1) improve beneficiary health through prevention; 2) educate, engage, and empower patients as members of their care team; and 3) provide accurate, timely, and actionable performance data to clinicians, patients, and other stakeholders.
Rating: Met

Feasibility	
eCQM feasibility testing/analysis conducted:	Yes, eCQM testing was performed [MERIT submission form; eCQM Feasibility Scorecard]
Feasibility: All data elements are in defined fields in electronic sources and align with USCDI/USCDI+ Quality standards making the measure highly feasible. The measure was tested in three EHRs and is highly feasible.	

Feasibility
The feasibility scorecard addresses the following domains: • Data availability: Data element exists in a structured format in this EHR. • Data accuracy: Information is from authoritative sources and/or is highly likely to be correct. • Data standards: Data element is coded in a nationally accepted terminology standard or can be mapped to that terminology standard. • Workflow: The data element is routinely collected during clinical care and requires no, or limited, additional data entry from a clinician or other provider, and no EHR interface changes. Feasibility testing identified two data elements that required additional review within the Oracle Health or Allscripts testing sites. For the BI-RADS result data elements, the feasibility plan indicated search terms will be specified for use in EHRs that do not capture BI-RADS in structured fields. The developer developed and validated the string search algorithm in Health Systems 1 and 2. During CBE endorsement review in 2024, the committee found the feasibility of this measure to be sufficiently demonstrated.
Rating: Met; Prior CBE Endorsement

Validity	
Validity testing method(s):	Face validity, patient-encounter level testing [MERIT Submission Form, Peer-reviewed journal article template]
Testing level(s):	Facility
Was this measure tested in the same target population as MIPS?	Yes
Validity: A technical expert panel (TEP) consisting of seven members, representing the patient experience and expertise in medicine, measure development, quality and safety of care, cancer screening, health services research, and EHRs, reviewed the measure. The majority of TEP members agreed that the measure can be used to distinguish good from poor quality care at the hospital (i.e., the facility) level. The developer conducted chart reviews in two health systems to validate the accuracy of eCQM automated patient allocations. Using stratified random samples, reviewers compared manual chart assessments—considered the gold standard—to eCQM results. Percentage agreements ranged from 97% to 99%, and Positive Predictive Values (PPVs) ranged from 99% to 100%, confirming the measure's strong validity. Health System 3 is currently undergoing similar validation. During CBE endorsement review in 2024, the committee found the validity of this measure to be sufficiently demonstrated. **Considerations for the committee:** The submitted face validity and patient-/encounter-level testing met the requirements for CBE endorsement. Committee members are encouraged to consider if this testing provides appropriate evidence of the measure's suitability for	

Validity
inclusion in MIPS.
Threats to validity: Threats to validity were not identified in the submission materials. Empiric validity testing was not completed for this measure.
Rating: Met; Prior CBE Endorsement

Reliability	
Reliability testing method(s):	Signal-to-Noise and Random Split-Half Correlation [MERIT Submission Form, Peer-reviewed journal article template]
Testing level:	Facility (Facility Group), Individual Clinician
Reliability discussion: Signal-to-noise reliability at the group level was calculated across six facility groups in one hospital system. The minimum reliability for the most recent year (2023) at the facility group level is 0.989 and 100% of the six facility groups have a reliability greater than 0.6. The developer calculated a Spearman's rank correlation between two randomly split halves of the data and reported a correlation of 0.94. In the original MERIT submission, ICC was calculated as the percentage of variation in facility-level scores attributable to facility-level signal variation, with 95% confidence intervals for each split sample. Battelle noted during review that this was not the type of ICC that measures correlation between the two split samples. These initial ICC values were very low: 0.019 for the test sample and 0.084 for the validation sample in 2020. These results conflicted with the signal-to-noise and Spearman rank results. In response to this issue noted during PA collaboration, the developer revised testing methods during the MERIT submission window to align with recommendations and calculated a different type of ICC to assess the correlation between the two spilt samples. The ICC calculation now aligns with the intended approach and is consistent with the Signal-to-Noise Ratio (SNR) and Spearman correlation results. These results are shown in Table 1, provided by the developer below. At the individual clinician level, the median SNR was 0.962 (95% CI: 0.917, 0.956) for the 99 clinicians at Health System 1. The minimum SNR was 0.142 and the maximum SNR was 0.989. The SNRs were >0.900 from 2020 to 2023 with relatively narrow 95% confidence intervals, indicating that a high proportion of overall variability is explained by the differences between measured entities (i.e., individual clinicians). The Spearman's rank correlation coefficient for 2023 at the individual level, was 0.79 (95% CI: 0.66, 0.87). Although substantially lower than at the facility group level, the overall clinician Spearman's rank correlation coefficient still indicated a strong positive correlation between the test and validation samples. During CBE endorsement review in 2024, the committee found the reliability of this measure to be sufficiently demonstrated.	
Additional reliability analyses: No additional reliability analyses were performed.	

Reliability
Rating: Met; Prior CBE Endorsement

Table 1. Intraclass Correlation Coefficients (ICC), Overall and by Year from 2018 to 2023 for Six Facility Groups in Health System 1

Measurement Year	Test-Validation Correlation	95% CI
Overall	0.996	(0.980, 0.999)
2018	0.929	(0.697, 0.987)
2019	0.835	(0.445, 0.970)
2020	0.904	(0.616, 0.982)
2021	0.942	(0.743, 0.989)

Usability	
Usability considered in application:	Yes, the submission materials briefly discuss the measure's usability within relevant programs.
Usability discussion: This measure has usability in MIPS. It has been successfully tested at the facility level and shown to be feasible for integration into EHR systems. Additionally, feedback from a patient representative on the TEP confirmed that reporting eCQM diagnostic rates would be meaningful and could positively influence patient decision-making. However, this measure is currently specified in Fast Healthcare Interoperability Resources (FHIR), which may result in implementation barriers within MIPS if program updates are delayed in future. During CBE endorsement review in 2024, the committee found the use/usability of this measure to be sufficiently demonstrated.	
Rating: Met, Prior CBE Endorsement	

Appropriateness of Scale

Appropriateness of Scale	
Similar or related measures in program(s):	Breast Cancer Screening Measure in MIPS Breast Cancer Screening Recall Rates in MIPS
Measure balance, burden, and value across target populations/measured entities: While there are similar measures within MIPS, this	

Appropriateness of Scale
measure assesses the full screening process from an inconclusive/abnormal screening mammogram through to diagnostic resolution, which offers benefit to the program population. The developer notes in submission materials that this eCQM emphasizes timeliness of diagnostic resolution after an abnormal screening mammogram to detect breast cancers and uses a patient-based—rather than episode-based—approach to measurement. The measure is patient based and complements two related measures that are already in use in CMS programs: the Breast Cancer Screening measure (CMIT ID: 00093) and the Breast Cancer Screening Recall Rates measure (CMIT ID: 01648). Additionally, this new measure does compete with one Healthcare Effectiveness Data and Information Set (HEDIS) measure: Follow-Up after Abnormal Breast Cancer Assessment. This competing measure reports on the percentage of mammograms with a BI-RADS of 0 that received follow-up diagnostic imaging within 90 days or mammograms with a BI-RADS of 4 or 5 that received follow-up breast biopsy within 90 days. However, the current competing measure does not quantify the percentage of patients who have timely diagnostic resolution from a screening mammogram to breast biopsy. Each of the related and competing clinical quality measures quantifies specific aspects of the multi-step breast cancer screening process; however, none of the measures assess the full screening process from an inconclusive/abnormal screening mammogram through to diagnostic resolution as the new measure does. Regarding balance of this measure's performance, burden and benefit across populations, the developer's literature review and analysis do not indicate a potential for differential benefit or harm to specific subgroups of participating entities or their patient populations. **Considerations for the committee:** Based on clinical and professional experience, the committee should consider the distribution of benefits and risks/burdens of the measure within the proposed program population.

Time-to-Value Realization

Time-to-Value Realization	
Plan for near- and long-term impacts after implementation:	None specified
Measure implementation impacts over time: The developer briefly mentions long- and near-term impacts of the measure as an eCQM in a patient population. There may be need for further examination of near- and long-term impacts of this measure after implementation across clinician and patient populations. **Considerations for the committee:** • What are the potential near- and long-term impacts of this measure on measured entities, proposed CMS program, and patient populations? • Will benefits and burdens associated with this measure be realized within an appropriate implementation time frame? • How will this measure mature through revisions in the future if added to the MIPS measure set?	