

MIPS Peer-Reviewed Journal Article Requirement Template

Section 101(c)(1) of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) requires submission of new measures for publication in applicable specialty-appropriate, peer-reviewed journals prior to implementing in the Merit-based Incentive Payment System (MIPS). Such measures will be submitted by the Centers for Medicare & Medicaid Services (CMS), to a journal(s), before including any new measure on the MIPS Quality Measures List. The measure submitter shall provide the required information for article submission under the MACRA per the MIPS Annual Call for Quality Measures submission process.

Interested parties submitting measures for consideration through the MIPS Annual Call for Quality Measures must complete the required information by the CMS Annual Call for Measures deadline (8 p.m. ET on May 1, 2025). Some of the information requested below may be listed in specific fields in the CMS Measures Under Consideration (MUC) Entry/Review Information Tool (MERIT); however, to ensure that CMS has all of the necessary information and avoid delays in the evaluation of your submission, please fully complete this form as an attached Word document. The information in MERIT must be consistent with the information below, including the following, but not limited to:

- **Measure Title:** Rate of Timely Follow-up on Abnormal Screening Mammograms for Breast Cancer Detection
- **Meaningful Measures 2.0 Framework Domain:** Closing Gaps of Care

Measure Steward: Brigham and Women's Hospital

Measure Developer: Brigham and Women's Hospital

Description: This electronic Clinical Quality Measure (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 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.

I. Statement

- Background (Why is this measure important?).

Breast cancer is the second most common cause of cancer deaths among women in the United States [1]. A study in early January 2024 predicted that around 42,250 women would die from breast cancer and an estimated 310,720 new cases of invasive breast cancer would be diagnosed that year [1].

Breast cancer survival is dependent upon cancer stage at diagnosis. Approximately 99% of women diagnosed with early-stage breast cancer live for five 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 (Breast Imaging – Reporting and Data System [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 Center for Disease Control and Prevention 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 [15]. In this cohort, women with no family history of breast cancer had lower follow-up rates, Black and Native American women had lower overall follow-up rates and lower biopsy rates [15]. 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 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.

References:

1. Key Statistics for Breast Cancer. American Cancer Society. Updated January 17, 2024. Accessed October 31, 2024. <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=bi-rads&source=search_result&selectedTitle=2%7E13&usage_type=default&display_rank=2#H24.
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.

- Environmental scan (Are there existing measures in this area?).

Related measures:

- Breast Cancer Screening (Higher rate = better): "Percentage of women 40-74 years of age who had a mammogram to screen for breast cancer in the 27 months prior to the end of the measurement period." (Quality ID #112, NQF #2372)
- Breast Cancer Screening Recall Rates (Target = 5 to 12%): "Percentage of beneficiaries with mammography or digital breast tomosynthesis (DBT) screening studies that are followed by a diagnostic mammography, DBT, ultrasound, or magnetic resonance imaging (MRI) of the breast in an outpatient or office setting within 45 days." (Hospital Outpatient Quality Reporting ID #1648)

Competing Healthcare Effectiveness Data and Information Set (HEDIS) measure:

- Follow-Up after Abnormal Breast Cancer Assessment (Higher rate = better): "The percentage of inconclusive or high-risk BI-RADS assessments that received appropriate follow-up within 90 days of the assessment, for members 40–74 years of age." (BCF-E). This 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; the measure does not quantify the percentage of patients that have timely diagnostic resolution from a screening mammogram to breast biopsy.

This novel eCQM emphasizes timeliness of diagnostic resolution after an abnormal screening mammogram to detect breast cancers. The measure is patient based, whereas the HEDIS measure targets health plans to report: "the percentage of inconclusive or high-risk BI-RADS assessments that received appropriate follow-up within 90 days of the assessment, for members 40–74 years of age." The HEDIS measure is based on mammography episodes, rather than patients. A patient may be included in the denominator multiple times in one measurement year, for example, if the patient received both an inconclusive screening mammogram and high-risk assessment on diagnostic follow-up imaging. The HEDIS measure is likely to face a ceiling effect by allowing for up to 180 days (90 days after inconclusive screening mammogram + 90 days after high-risk assessment on diagnostic follow-up imaging) to diagnostic resolution requiring biopsy.

The HEDIS measure includes only diagnostic mammograms and breast ultrasounds as appropriate follow-up after an inconclusive (BI-RADS 0) assessment. Our eCQM is more comprehensive by also including breast MRI as appropriate diagnostic follow-up.

Each of the related and competing clinical quality measures quantifies specific aspects of the multi-step breast cancer screening process; however, the proposed eCQM assesses the full screening process from an inconclusive/abnormal screening mammogram through to diagnostic resolution.

II. Gap Analysis

- Provide evidence for the measure (What are the gaps and opportunities to improve care?).
- Expected outcome (patient care/patient health improvements, cost savings).
- Recommendation for the measure (Is it based on a study, consensus opinion, USPSTF recommendation etc.?).

Despite advancements in therapies for breast cancer, early detection via routine mammographic screening has had a substantial impact in reducing breast cancer mortality since the 1990s [1]. Breast cancer has the highest treatment cost of any cancer, costing over \$26.2 billion for medical services and \$3.5 billion for prescription medications [2]. Early detection through screening can reduce treatment costs by 30-100% [3]. The National Breast and Cervical Cancer Early Detection Program (NBCCEDP) is estimated to have saved 369,000 life-years compared to no screening for the 1.8 million women screened in the program between 1991-2006 [4].

The United States Preventive Services Task Force (USPSTF) guidelines recommend universal breast cancer screening in patients aged 40-74 years (Grade B recommendation) [5]. The Center for Disease Control and Prevention National Breast and Cervical Cancer Early Detection Program (NBCCEDP) recommends that breast cancer screening to diagnostic resolution occur within 60 days [6]. A systematic review by Doubeni et al. (2018) concluded that “the current published literature does not confirm or refute the 60-day threshold set by the BCCEDP. Given evidence of an increase in risk after 90 days, the 60-day goal set in the Centers for Disease Control and Prevention quality guidelines is a reasonable target for quality improvement after a positive breast cancer screening” [7]. Delays in diagnostic follow-up after abnormal screening results worsen prognostic outcomes for patients by prolonging the initiation of early, lower-cost, and less invasive interventions at diagnosis [1]. In particular, invasive triple negative breast cancers account for 10-15% of all newly diagnosed breast cancers in the United States [8]. These tumors can double in size in <60 days and spread rapidly [9]. However, the type of breast cancer can only be confirmed through biopsy, making timely access to screening and diagnostic evaluation critical, so that women with invasive cancers are diagnosed and treated prior to metastasis and when therapies are more likely to be effective.

Certain racial and ethnic groups, including Black, Asian, and Hispanic women, as well as lower-income patients and those living in rural areas have lower rates of follow-up and timely diagnostic resolution, increasing their risks of later-stage diagnoses and death [10, 11, 12]. These differences in breast cancer outcomes emphasize the need for robust systems and protocols that monitor delays in follow-up to ensure access to timely and complete care after all abnormal screening mammogram results.

Currently, federal requirements instituted by the Mammography Quality Standards Act are limited in their capacity to improve quality performance in breast imaging facilities because they do not provide guidelines on how to assess diagnostic timeliness [14, 15]. There is also no requirement for facilities to track and report abnormal screening mammogram (recall) rates and early cancer detection rates which can further inform the quality of mammography outcomes and practices in breast imaging facilities [13].

A 2014 study found that breast imaging facilities are able to leverage routinely collected data to measure their ability to meet certain breast cancer screening diagnostic quality benchmarks [14]. These benchmarks included timely follow-up imaging (more than 90% of patients should receive diagnostic imaging within 30 days of an abnormal screening mammogram) and timely biopsy (more than 90% of patients should receive a recommended biopsy within 60 days of an abnormal screening mammogram) [14]. Only 62% of participating facilities (n=52) were able to show that they met the benchmark for timely follow-up imaging and 27% met the benchmark for timely diagnostic biopsy, highlighting the need to improve facility performance on the timeliness of diagnostic imaging and biopsy [14].

A follow-up 2021 study by the same team found that facilities were also significantly more likely to reach these mammography quality benchmarks for timeliness of follow-up the longer they participated in a quality measurement program [15]. In this study, facilities not designated as an American College of Radiology (ACR) Breast Imaging Center of Excellence showed the most improvement in recall rate, proportion not lost to follow-up at imaging, biopsy recommendation rate, and early-stage cancer detection [15, 16]. Therefore,

comprehensive quality improvement initiatives with stated performance benchmarks positively affect mammography practices and outcomes, with greater impact on previously underperforming facilities.

Results of regular quality assessments of breast cancer screening and timely diagnostic follow-up can be used to support the implementation of interventions, such as patient navigation and case management, electronic health record (EHR) reminders, and patient education and outreach, that have been shown to improve follow-up rates [12, 17, 18].

This electronic clinical quality measure (eCQM) uses standard terminologies to calculate the rate of timely diagnostic resolution in facilities that perform mammographic screening and follow-up. The specifications of this measure are supported by findings outlined in peer-reviewed literature, current screening guidelines, and existing related clinical quality measures. Facilities can use this tool to conduct routine quality assessment checks to guide quality improvement initiatives aiming to address timeliness of care in the breast cancer screening and diagnostic processes.

References:

1. 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.
2. National Cancer Institute. Financial burden of cancer care. Cancer Trends Progress Report. Reviewed March 2024. Accessed October 2024. https://progressreport.cancer.gov/after/economic_burden.
3. Feig S. Cost-effectiveness of mammography, MRI, and ultrasonography for breast cancer screening. *Radiol Clin North Am*. 2010;48(5):879-891. doi:10.1016/j.rcl.2010.06.002. PMID: 20868891.
4. Hoerger TJ, Ekwueme DU, Miller JW et al. (2011) Estimated effects of the National Breast and Cervical Cancer Early Detection Program on breast cancer mortality. *Am J Prev Med* 40:397–404. PMID: 21406272.
5. US Preventive Services Task Force; Nicholson WK, Silverstein M, Wong JB, Barry MJ, Chelmow D, Coker TR, Davis EM, Jaén CR, Krousel-Wood M, Lee S, Li L, Mangione CM, Rao G, Ruiz JM, Stevermer JJ, Tsevat J, Underwood SM, Wiehe S. Screening for Breast Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2024 Jun 11;331(22):1918-1930. doi: 10.1001/jama.2024.5534. Erratum in: *JAMA*. 2024 Sep 30. doi: 10.1001/jama.2024.19851. PMID: 38687503.
6. 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.
7. Doubeni CA, Gabler NB, Wheeler CM, McCarthy AM, Castle PE, Halm EA, Schnall MD, Skinner CS, Tosteson ANA, Weaver DL, Vachani A, Mehta SJ, Rendle KA, Fedewa SA, Corley DA, Armstrong K. Timely follow-up of positive cancer screening results: A systematic review and recommendations from the PROSPR Consortium. *CA Cancer J Clin*. 2018 May;68(3):199-216. doi: 10.3322/caac.21452. Epub 2018 Mar 30. PMID: 29603147; PMCID: PMC5980732.
8. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. *Cancer*. 2007 May 1;109(9):1721-8. doi: 10.1002/cncr.22618. PMID: 17387718.
9. 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.
10. 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.
11. Nguyen KH, Pasick RJ, Stewart SL, Kerlikowske K, Karliner LS. Disparities in abnormal mammogram follow-up time for Asian women compared with non-Hispanic white women and between Asian ethnic groups. *Cancer*. 2017;123(18):3468-3475. doi:10.1002/cncr.30756. PMID: 28603859.
12. 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.

13. Richardson LC, Henley SJ, Miller JW, Massetti G, Thomas CC. Patterns and Trends in Age-Specific Black-White Differences in Breast Cancer Incidence and Mortality - United States, 1999-2014. *MMWR Morb Mortal Wkly Rep.* 2016;65(40):1093-1098. Published 2016 Oct 14. doi:10.15585/mmwr.mm6540a1. PMID: 27736827.
14. Rauscher GH, Murphy AM, Orsi JM, Dupuy DM, Grabler PM, Weldon CB. Beyond the mammography quality standards act: measuring the quality of breast cancer screening programs. *AJR Am J Roentgenol.* 2014;202(1):145-151. doi:10.2214/AJR.13.10806. PMID: 24261339.
15. Rauscher GH, Tossas-Milligan K, Macarol T, Grabler PM, Murphy AM. Trends in Attaining Mammography Quality Benchmarks With Repeated Participation in a Quality Measurement Program: Going Beyond the Mammography Quality Standards Act to Address Breast Cancer Disparities. *J Am Coll Radiol.* 2020;17(11):1420-1428. doi:10.1016/j.jacr.2020.07.019. PMID: 32771493.
16. Albus K. ACR Designated Comprehensive Breast Imaging Center (CBIC)(Revised 9-29-23). American College of Radiology. September 9, 2023. Accessed October 31, 2024. <https://accreditationsupport.acr.org/support/solutions/articles/11000068075-acr-designated-comprehensive-breast-imaging-center-cbic-revised-9-29-23->.
17. Haas JS, Atlas SJ, Wright A, et al. Multilevel Follow-up of Cancer Screening (mFOCUS): Protocol for a multilevel intervention to improve the follow-up of abnormal cancer screening test results. *Contemp Clin Trials.* 2021;109:106533. doi:10.1016/j.cct.2021.106533. PMID: 34375748.
18. Atlas SJ, Tosteson ANA, Wright A, et al. A Multilevel Primary Care Intervention to Improve Follow-Up of Overdue Abnormal Cancer Screening Test Results: A Cluster Randomized Clinical Trial. *JAMA.* 2023;330(14):1348-1358. doi:10.1001/jama.2023.18755. PMID: 37815566.

III. Reliability/Validity

- What testing has been performed at the level of implementation? (MIPS requires full measure testing at the individual clinician level (and may also need to be tested at the group level) for MIPS Clinical Quality Measures (CQMs) and Electronic Clinical Quality Measures (eCQMs) collection types. Administrative claims measures tested at the group level require a reliability threshold to be implemented at the group level.)

Please provide testing results including the N value, Bonnie test case results, correlation coefficient and any other pertinent information or values to be considered.

Measure Specifications

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.

1. 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].
2. 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].
3. 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].
4. Patients that received negative/benign/probably benign follow-up imaging or breast biopsy within 60 days are included in the numerator population.

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.

1. 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].
2. 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].
3. 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].
4. Patients with at least one abnormal screening or screening-to-diagnostic mammogram are included in the denominator population.

MADiE results: 40 test cases passed with 100% coverage.

Evidence of Performance Gap

Table 1 presents health system characteristics and overall eCQM rates, and **Table 2** shows the patient and test characteristics for the included samples by health system.

Table 1. Health System Characteristics and Overall eCQM Rates by Health System

| Characteristics | Health System 1 | Health System 2 | | Health System 3 |
|--|---------------------------|-----------------|---------------|-----------------|
| BI-RADS data fields | Structured & Unstructured | Structured | Unstructured | Unstructured |
| Screening mammograms — no. | 553,183 | 32,004 | 129,937 | 91,915 |
| Abnormal screening mammograms — no. | 76,719 | Not reported | 13,228 | 7,676 |
| Index abnormal screening mammograms — no. | 76,598 | 3,318 | 13,213 | 7,625 |
| Number with timely diagnostic resolution — no. | 70,304 | 2,945 | 12,322 | 6,897 |
| eCQM rate of timely diagnostic resolution — % | 91.8% | 88.8% | 93.3% | 90.5% |
| Number of facility groups — no. | 6 | 1 | 5 | 1 |
| Number of measurement years — no. (range of years) | 6 (2018-2023) | 6 (2018-2023) | 5 (2019-2023) | 7 (2016-2023) |

Table 2. Patient Characteristics for Included Samples by Health System

| Characteristics | Health System 1 (n = 76,598) | Health System 2 – Structured (n = 3,318) | Health System 2 – Unstructured (n = 13,213) | Health System 3 – (n = 7,625) |
|--|---------------------------------|---|--|----------------------------------|
| Age at index abnormal screening mammogram | | | | |
| Mean (SD) — yr | 55.2 (10.1) | 56.4 (9.8) | 56.0 (9.9) | 53.0 (9.9) |

| | | | | |
|--|---------------|--------------|---------------|--------------|
| Distribution — no. (%) | | | | |
| 40 to 49 yr | 26,889 (35.1) | 962 (29.0) | 4,251 (32.2) | 3,019 (39.6) |
| 50 to 64 yr | 32,614 (42.5) | 1,542 (46.5) | 5,942 (45.0) | 3,179 (41.7) |
| 65 to 75 yr | 17,095 (22.3) | 814 (24.5) | 3,020 (22.9) | 1,427 (18.7) |
| Sex — no. (%) | | | | |
| Female | 76,598 (100) | 3,318 (100) | 13,213 (100) | 7,625 (100) |
| Race — no. (%) | | | | |
| Asian | 4,015 (5.2) | 29 (0.9) | 287 (2.2) | 259 (3.4) |
| Black | 4,656 (6.1) | 186 (5.6) | 671 (5.1) | 405 (5.3) |
| White | 61,254 (80.0) | 2,097 (63.2) | 9,701 (73.4) | 6,197 (81.3) |
| Other* | 211 (0.3) | 997 (30.0) | 2,451 (18.5) | 119 (1.6) |
| Unknown | 6,462 (8.4) | 9 (0.3) | 104 (0.79) | 59 (0.8) |
| Ethnic group — no. (%) | | | | |
| Hispanic | 6,570 (8.6) | 937 (28.2) | 1,872 (14.2) | 335 (4.4) |
| Non-Hispanic | 63,655 (83.1) | 2,304 (69.4) | 10,945 (82.8) | 7,368 (96.6) |
| Unknown | 6,373 (8.3) | 77 (2.3) | 396 (3.0) | 257 (3.4) |
| Primary insurance at index abnormal screening mammogram — no. (%) | | | | |
| Private | 56,507 (73.8) | 2,183 (65.8) | Not reported | 2,006 (26.3) |
| Medicare | 15,126 (19.9) | 720 (21.7) | Not reported | 1,507 (19.8) |
| Medicaid | 1,865 (2.4) | 349 (10.5) | Not reported | 788 (10.3) |
| Other | 183 (0.2) | 59 (1.8) | Not reported | 332 (4.4) |
| Unknown | 2,917 (3.8) | 7 (0.2) | Not reported | 2,992 (39.2) |
| Primary language — no. (%) | | | | |
| English | 71,005 (92.7) | 2,728 (82.2) | 11,856 (89.7) | 6,946 (91.1) |
| Spanish | 3,349 (4.37) | 571 (17.2) | 1,110 (8.4) | 486 (6.4) |
| Other | 1,702 (2.2) | 10 (0.3) | 168 (1.3) | 176 (2.3) |
| Unknown | 542 (0.7) | 9 (0.3) | 79 (0.60) | 17 (0.2) |

*Patients identified as American Indian, Alaska Native, or Native Hawaiian or Other Pacific Islander.

Tables 3 and 4 present eCQM rates for Health System 1 at the integrated delivery system and hospital-affiliated facility group levels, respectively. Rates for Health System 2 calculated based on data extracted from structured BI-RADS fields at one facility group are presented in **Table 5**, and rates calculated based on unstructured BI-RADS data for 5 facility groups are provided in **Tables 6 and 7**. **Table 8** shows the eCQM rates for one facility group from Health System 3.

The overall eCQM performance rates for the three healthcare systems were high at 91.9% (95% CI: 91.6, 92.0), 88.8% (95% CI: 87.6, 89.8), 93.3% (95% CI: 92.8, 93.7), and 90.5% (95% CI: 89.8, 91.1). The eCQM facility group performance rates across all three health systems and years ranged from 62.0% (95% CI: 58.6, 65.3) to 97.6% (95% CI: 96.0, 98.5), demonstrating large significant differences across facility groups and years, as well as substantial opportunities for improvement for some facility groups.

These differences were most pronounced in 2023, which is most reflective of current performance, in Health System 1. The annual eCQM performance rate at the integrated delivery system level for 2023 was significantly lower at 84.4% (95% CI: 83.7, 85.0) than for 2022, which had an annual eCQM performance rate of 91.6% (95% CI: 91.1, 92.0). The facility group level eCQM rates show that the performance significantly decreased in three facility groups in 2023 to 62.0% (95% CI: 58.6, 65.3), 75.2% (95% CI: 73.8, 76.6), and 85.1% (95% CI: 83.1, 87.1), indicating where efforts to improve follow-up should be focused.

These results highlight the importance of tracking performance over time even in health systems and facility groups that have been consistently performing well.

Table 3. eCQM Rates by Year for the Integrated Delivery System for **Health System 1**

| Measurement Year | eCQM Rate of Timely Diagnostic Resolution (95% CI) |
|------------------|--|
| Overall | 91.9 (91.6, 92.0) |
| 2018 | 94.2 (93.7, 94.6) |

| | |
|------|-------------------|
| 2019 | 93.6 (93.2, 94.0) |
| 2020 | 93.1 (92.6, 93.6) |
| 2021 | 94.0 (93.6, 94.3) |
| 2022 | 91.6 (91.1, 92.0) |
| 2023 | 84.4 (83.7, 85.0) |

Table 4. eCQM Rates by Year for 6 Hospital-affiliated Facility Groups at **Health System 1**

| Measurement Year | eCQM Rate of Timely Diagnostic Resolution (95% CI) | | | | | |
|------------------|--|------------------------------------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| | Facility Group 1 | Facility Group 2 | Facility Group 3 | Facility Group 4 | Facility Group 5 | Facility Group 6 |
| Overall | 86.6 (85.6, 87.7) | 87.2 (86.7, 87.7) | 91.3 (90.7, 92.0) | 93.5 (93.1, 93.9) | 94.7 (93.8, 94.7) | 95.3 (95.0, 95.6) |
| 2018 | 96.7 (95.4, 98.0) | 92.4 (91.4, 93.6) | 93.7 (92.6, 94.9) | 91.6 (90.6, 92.6) | 95.7 (94.8, 96.7) | 97.2 (96.5, 97.8) |
| 2019 | 94.9 (93.2, 96.7) | 92.8 (91.8, 93.7) | 93.8 (92.5, 95.1) | 91.6 (90.7, 92.5) | 93.6 (92.6, 94.7) | 96.6 (95.9, 97.2) |
| 2020 | 92.5 (90.1, 94.8) | 89.2 (88.0, 90.4) | 93.0 (91.5, 94.5) | 92.6 (91.5, 93.6) | 95.5 (94.4, 96.6) | 97.6 (96.0, 98.5) |
| 2021 | 93.5 (91.7, 95.4) | 90.7 (89.8, 91.6) | 90.8 (89.4, 92.3) | 96.4 (95.7, 97.1) | 94.2 (93.2, 95.1) | 97.3 (96.7, 97.8) |
| 2022 | 86.8 (84.6, 89.0) | 86.0 (84.9, 87.1) | 91.1 (89.6, 92.6) | 95.4 (94.6, 96.2) | 94.7 (93.6, 95.7) | 95.1 (94.3, 95.8) |
| 2023 | 62.0 (58.6, 65.3) | 75.2 (73.8, 76.6) | 85.1 (83.1, 87.1) | 94.0 (93.1, 94.9) | 91.8 (90.3, 93.3) | 89.1 (88.0, 90.2) |

Table 5. eCQM Rates by Year for 1 Facility Group at **Health System 2** (structured BI-RADS data)

| Measurement Year | eCQM Rate of Timely Diagnostic Resolution (95% CI) |
|------------------|--|
| Overall | 88.8 (87.6, 89.8) |
| 2018 | 81.9 (76.7, 86.3) |
| 2019 | 85.0 (81.8, 87.9) |
| 2020 | 90.6 (87.7, 93.0) |
| 2021 | 91.7 (89.3, 93.8) |
| 2022 | 88.5 (85.7, 90.9) |
| 2023 | 90.6 (88.3, 92.6) |

Table 6. eCQM Rates by Year for 5 Facility Groups at **Health System 2** (unstructured BI-RADS data)

| Measurement Year | eCQM Rate of Timely Diagnostic Resolution (95% CI) |
|------------------|--|
| Overall | 93.3 (92.8, 93.7) |
| 2019 | 92.5 (91.4, 93.4) |
| 2020 | 93.1 (92.0, 94.2) |
| 2021 | 94.5 (93.6, 95.3) |
| 2022 | 93.1 (92.1, 94.0) |
| 2023 | 93.0 (92.1, 93.8) |

Table 7. eCQM Rates by Facility Group for 5 Facility Groups at **Health System 2** (unstructured BI-RADS data)

| Facility Group | eCQM Rate of Timely Diagnostic Resolution (95% CI) |
|----------------|--|
| Overall | 93.3 (92.8, 93.7) |
| 1 | 96.8 (94.9, 98.0) |
| 2 | 95.8 (95.3, 96.3) |
| 3 | 95.0 (90.1, 97.6) |
| 4 | 94.8 (94.8, 97.7) |
| 5 | 88.7 (87.8, 89.5) |

Table 8. eCQM Rates by Year for 1 Hospital-affiliated Facility Group at **Health System 3**

| Measurement Year | eCQM Rate of Timely Diagnostic Resolution (95% CI) |
|------------------|--|
| Overall | 90.5 (89.8, 91.1) |
| 2018 | 91.3 (89.1, 93.1) |
| 2019 | 91.3 (89.0, 93.1) |
| 2020 | 92.0 (89.7, 93.8) |
| 2021 | 94.0 (92.1, 95.5) |
| 2022 | 89.6 (87.8, 91.2) |
| 2023 | 89.0 (87.4, 90.5) |

Table 15 (Appendix 1) presents overall eCQM rates for individual clinicians at Health System 1, which ranged from 70.4% (95% CI: 59.8, 81.0) to 98.7% (95% CI: 97.2, 99.9). Rates were reported for the 99 (of 111) individual clinicians included in the reliability analyses, which required a denominator of at least 40 index abnormal screening mammograms across all years (2018-2023).

The Technical Expert Panel (TEP) supported the follow-up timeframe of 60 days based on standards defined by the Center for Disease Control and Prevention National Breast and Cervical Cancer Early Detection Program. This timeframe was also supported by the published literature which demonstrated that invasive triple negative breast cancers have been shown to double in size in less than 60 days. The TEP and other stakeholders recommended shortening the timeframe in future iterations of the eCQM specifications as the timeliness of care improves. In response, the eCQM rates were also calculated for 30-day (Appendix 2) and 45-day (Appendix 3) follow-up timeframes.

- Reliability Testing Results at the accountable entity level

Signal-to-Noise Analyses

Signal-to-Noise Ratios (SNR) were calculated for the six hospital-affiliated facility groups and 99 individual clinicians (i.e., breast radiologists) at Health System 1. The signal-to-noise analyses estimated the proportion of overall variability explained by the differences between measured entities. A minimum sample size of 40 patients was required for the signal-to-noise analysis. The results are reported overall and by year from 2018 to 2023, since the measure is intended to be reported annually.

The facility group SNRs are provided in **Table 9**. Overall, the median SNR was 0.996 (95% CI: 0.988, 0.998) for the six hospital-affiliated facility groups at Health System 1. The minimum SNR was 0.914 and the maximum SNR was 0.999. The SNRs were high across all years from 2018 to 2023. The median SNR for 2023, which is most reflective of current performance, was 0.997 (95% CI: 0.991, 0.999) for the six hospital-affiliated facility groups. The minimum SNR was 0.989 and the maximum SNR was 0.998 in 2023.

The >0.97 median SNRs with narrow 95% confidence intervals, overall and across all years, indicate that a very high proportion of overall variability is explained by the differences between measured entities (i.e., hospital-affiliated facility groups).

Table 9. Signal-to-Noise Ratios (SNR), Overall and by Year from 2018 to 2023 for Six Facility Groups in **Health System 1**

| Measurement Year | Median SNR (95% CI) | Minimum SNR | Maximum SNR |
|------------------|-----------------------------|--------------|--------------|
| Overall | 0.996 (0.988, 0.998) | 0.914 | 0.999 |
| 2018 | 0.992 (0.973, 0.999) | 0.968 | 0.993 |
| 2019 | 0.981 (0.930, 0.999) | 0.914 | 0.985 |
| 2020 | 0.991 (0.969, 0.999) | 0.962 | 0.994 |
| 2021 | 0.974 (0.993, 0.999) | 0.968 | 0.995 |
| 2022 | 0.994 (0.984, 0.999) | 0.981 | 0.996 |
| 2023 | 0.997 (0.991, 0.999) | 0.989 | 0.998 |

The individual clinician SNRs are provided in **Table 10**. Overall, 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).

Table 10. Signal-to-Noise Ratios (SNR), Overall and by Year from 2018 to 2023 for 99 Individual Clinicians in Health System 1

| Measurement Year | Median SNR (95% CI) | Minimum SNR | Maximum SNR |
|------------------|-----------------------------|--------------|--------------|
| Overall | 0.962 (0.917, 0.956) | 0.142 | 0.989 |
| 2018 | 0.880 (0.823, 0.886) | 0.103 | 0.955 |
| 2019 | 0.810 (0.693, 0.786) | 0.040 | 0.931 |
| 2020 | 0.920 (0.847, 0.920) | 0.132 | 0.955 |
| 2021 | 0.913 (0.875, 0.908) | 0.667 | 0.953 |
| 2022 | 0.932 (0.880, 0.940) | 0.182 | 0.960 |
| 2023 | 0.956 (0.869, 0.950) | 0.165 | 0.989 |

Random Split-Half Correlation Analyses

A random half split correlation was conducted at the hospital-affiliated facility group and individual clinician levels at Health System 1, with 6 facility groups and 99 individual clinicians included in the analyses. To perform a random half split correlation analysis, we required a minimum of 40 patients for each facility group or individual clinician per year (20 patients in each split sample). Patients were randomly split by facility group or individual clinician and year into a test sample or a validation sample, with ~50% of patients in each sample. The descriptive statistics and p-values for each group were calculated. Spearman's rank correlation coefficients and Interclass Correlation Coefficients (ICC) were calculated with 95% confidence intervals. The ICCs were calculated to describe how much variation in the facility group or individual clinician level scores was due to facility group or clinician level signal variation. The Spearman's rank correlation coefficients were calculated to compare the relative rankings of facility groups or clinicians in the test and validation samples. The Spearman's rank correlation coefficients and ICCs were reported overall and by year since the measure is intended to be reported annually.

Spearman's Rank Correlation Coefficient

The overall facility group Spearman's rank correlation coefficient was 1.00 (95% CI: 0.99, 0.99) (**Table 11**). There was no apparent trend over time. The Spearman's rank correlation coefficient for 2023, which is most reflective of current performance, was 0.94 (95% CI: 0.49, 0.99). The overall Spearman's rank correlation coefficient of 1.00 (95% CI: 0.99, 0.99) indicated a very strong positive correlation between the test and validation samples. However, confidence intervals by year were wide given that only six facility groups were included in the analysis. Additional facility group data is required to generate narrower confidence intervals.

Table 11. Spearman's Rank Correlation Coefficients, Overall and by Year from 2018 to 2023 for Six Facility Groups in Health System 1

| Measurement Year | Test-Validation Correlation | 95% CI |
|------------------|-----------------------------|---------------------|
| Overall | 1.00 | (0.99, 0.99) |
| 2018 | 1.00 | (0.99, 0.99) |
| 2019 | 0.94 | (0.49, 0.99) |
| 2020 | 0.77 | (-0.18, 0.97) |
| 2021 | 0.94 | (0.49, 0.99) |
| 2022 | 0.83 | (-0.03, 0.98) |
| 2023 | 0.94 | (0.49, 0.99) |

The overall individual clinician Spearman's rank correlation coefficient was 0.80 (95% CI: 0.72, 0.86) (**Table 12**). There was no apparent trend over time. The Spearman's rank correlation coefficient for 2023, which is most reflective of current performance, 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.

Table 12. Spearman's Rank Correlation Coefficients, Overall and by Year from 2018 to 2023 for 99 Individual Clinicians in **Health System 1**

| Measurement Year | Test-Validation Correlation | 95% CI |
|------------------|-----------------------------|---------------------|
| Overall | 0.80 | (0.72, 0.86) |
| 2018 | 0.68 | (0.51, 0.80) |
| 2019 | 0.30 | (0.05, 0.51) |
| 2020 | 0.64 | (0.45, 0.77) |
| 2021 | 0.80 | (0.72, 0.86) |
| 2022 | 0.56 | (0.34, 0.72) |
| 2023 | 0.79 | (0.66, 0.87) |

Interclass Correlation Coefficients (ICC)

The overall facility group level ICC was 0.037 (95% CI: 0.013, 0.302) in the test sample and 0.066 (95% CI: 0.027, 0.306) in the validation sample (**Table 13**). There were no apparent trends over time. The overall ICCs indicated that a low proportion of variation in the facility group level scores was due to facility group level signal variation. Notably, the 95% CIs were very wide given that only six facility groups were included in the analysis. Additional facility group data is required to generate narrower confidence intervals.

Table 13. Interclass Correlation Coefficients (ICC), Overall and by Year from 2018 to 2023 for Six Facility Groups in **Health System 1**

| Measurement Year | Test Sample (95% CI) | Validation Sample (95% CI) |
|------------------|-----------------------------|-----------------------------|
| Overall | 0.037 (0.013, 0.302) | 0.066 (0.027, 0.306) |
| 2018 | 0.002 (0.001, 0.999) | 0.066 (0.025, 0.366) |
| 2019 | 0.020 (0.004, 0.914) | 0.032 (0.011, 0.236) |
| 2020 | 0.019 (0.003, 0.999) | 0.084 (0.033, 0.404) |
| 2021 | 0.052 (0.012, 0.909) | 0.100 (0.040, 0.433) |
| 2022 | 0.007 (0.001, 0.999) | 0.097 (0.039, 0.417) |
| 2023 | 0.019 (0.003, 0.999) | 0.084 (0.033, 0.404) |

The overall individual clinician ICC was 0.091 (95% CI: 0.068, 0.130) in the test sample and 0.093 (95% CI: 0.069, 0.131) in the validation sample (**Table 14**). There were no apparent trends over time. The overall ICCs indicated that a low proportion of variation in the individual clinician level scores was due to individual clinician level signal variation.

Table 14. Interclass Correlation Coefficients (ICC), Overall and by Year from 2018 to 2023 for 99 Individual Clinicians in **Health System 1**

| Measurement Year | Test Sample (95% CI) | Validation Sample (95% CI) |
|------------------|-----------------------------|-----------------------------|
| Overall | 0.091 (0.068, 0.130) | 0.093 (0.069, 0.131) |
| 2018 | 0.034 (0.017, 0.097) | 0.061 (0.034, 0.137) |
| 2019 | 0.026 (0.012, 0.087) | 0.040 (0.022, 0.093) |
| 2020 | 0.092 (0.055, 0.180) | 0.076 (0.045, 0.155) |
| 2021 | 0.074 (0.045, 0.142) | 0.054 (0.031, 0.112) |
| 2022 | 0.086 (0.055, 0.151) | 0.090 (0.058, 0.158) |
| 2023 | 0.099 (0.067, 0.160) | 0.148 (0.103, 0.226) |

- Face Validity Testing Results, Clinician Sites

Technical Expert Panel (TEP) Face Validity Voting

The Technical Expert Panel (TEP) consisted of six members, representing the patient experience and expertise in medicine, measure development, quality and safety of care, cancer screening, health services research, and electronic health records (EHRs). During a July 2024 meeting, the TEP was presented with final

measure specifications and revised rate calculations at the integrated health system level and the hospital (i.e., hospital-affiliated facility group level). The TEP also had an opportunity to discuss questions and provide feedback to the measure development team at this time. A formal face validity vote was conducted using the polling function in Zoom.

The TEP members were asked to agree (vote YES) or disagree (vote NO) on the following two statements:

1. The Timely Follow-up on Abnormal Screening Mammography for Breast Cancer Detection – eCQM, as specified at the integrated health system level, can be used to distinguish good from poor quality care.
2. The Timely Follow-up on Abnormal Screening Mammography for Breast Cancer Detection – eCQM, as specified at the hospital level, can be used to distinguish good from poor quality care.

The final vote for #1 was 6/6 members (100%) in agreement with the statement at the integrated health system level. The final vote for #2 was 5/6 members (83.3%) in agreement with the statement at the hospital level.

The majority of TEP members agreed that the measure can be used to distinguish good from poor quality care at the hospital (i.e., facility group) level.

- Empiric Validity Testing Results at the accountable entity level

Empiric validity testing was not conducted at the accountable-entity level.

- Data Element/Patient Encounter Level Testing

Patient-level Data Element Validity Percentage Agreement and Positive Predictive Value

Chart reviews were conducted on the final data extractions using stratified random samples of 100 patients at Health System 1 and 200 patients at Health System 2 (100 for structured BI-RADS fields and 100 for unstructured BI-RADS fields) to assess whether the eCQM appropriately allocated patients into the numerator or denominator only to calculate the eCQM performance rates. Given the high overall eCQM rates at Health Systems 1 and 2, patients that did not meet the numerator criteria as assigned by the eCQM were oversampled; 80 patients that met the numerator criteria and 20 patients that did not meet the numerator criteria were selected for chart review. Manual chart review was considered the criterion standard. Chart reviewers were blinded to the eCQM automated allocations and reviewed the full chart to assess whether each patient should be included in the numerator or denominator only. Percentage agreement was calculated between the criterion-standard manual chart review allocations and the eCQM automated allocations. The Positive Predictive Value (PPV) of the denominator was also calculated to quantify the proportion of patients included in the denominator that required immediate follow-up with diagnostic breast imaging or breast biopsy. Health System 3 is in the process of conducting chart reviews.

Health System 1

Percentage agreement: 99%

PPV: 99%

A facility at Health System 1 documented BI-RADS data in unstructured fields from 2018-2022. To address this issue, a string search was developed to extract BI-RADS results from initial reports and addenda. The algorithm showed 100% accuracy for BI-RADS category extraction for the last overall assessment from a random sample of 100 breast imaging reports based on manual chart reviews.

Health System 2 – Structured BI-RADS Data

Percentage agreement: 97%

PPV: 100%

Health System 2 – Unstructured BI-RADS Data

Percentage agreement: 98%

PPV: 100%

The near-perfect percentage agreements and PPVs demonstrated strong validity of the eCQM automated allocations and ability to calculate accurate eCQM rates.

- Exclusion Frequency

This eCQM does not have any denominator exclusions.

- What were the minimum sample sizes used for reliability results?

A minimum of 40 patients per group (i.e., facility group or individual clinician) per year (20 patients in each split sample) were required for reliability testing.

- Other Information
 - Is it risk adjusted? If so, how?
The eCQM is not risk adjusted.

- What benchmarking information is available?

Three studies identified through the environmental scan of the literature provided benchmarking information:

| Metric | Description | Benchmark |
|--|--|-----------|
| Diagnostic follow-up completion [1] | Completion of diagnostic follow-up after an abnormal screening mammography result | ≥90% |
| Timely diagnostic follow-up completion [1] | Percentage of patients where diagnostic follow-up was completed within >60 days after an abnormal screening mammography result | ≤25% |
| Timely follow-up imaging [2,3] | Completion of diagnostic imaging within ≤30 days of an abnormal screening mammography result | ≥90% |
| Timely biopsy [2,3] | Completion of biopsy within ≤60 days of an abnormal screening mammography result | ≥90% |

References:

1. 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. <https://doi.org/10.1002/cncr.28817>. PMID: 25099899.
2. Rauscher GH, Murphy AM, Orsi JM, Dupuy DM, Grabler PM, Weldon CB. Beyond the mammography quality standards act: measuring the quality of breast cancer screening programs. *AJR Am J Roentgenol*. 2014;202(1):145-151. <https://doi.org/10.2214/AJR.13.10806>. PMID: 24261339.
3. Rauscher GH, Tossas-Milligan K, Macarol T, Grabler PM, Murphy AM. Trends in Attaining Mammography Quality Benchmarks With Repeated Participation in a Quality Measurement Program: Going Beyond the Mammography Quality Standards Act to Address Breast Cancer Disparities. *J Am Coll Radiol*. 2020;17(11):1420-1428. <https://doi.org/10.1016/j.jacr.2020.07.019>. PMID: 32771493.

- Collection Type: Specify the data collection type.
This eCQM leverages electronic health record (EHR) data.
- Specify measure stage of development.

Completed stages of development: measure conceptualization, measure specification, and measure testing.

- For Patient Reported Outcome Performance Measures:
 - The survey or tool has been tested and doesn't require modifications based on results?
 - Patient/encounter level testing for each critical data element doesn't require changes to the tool base on the results?

This eCQM is not a Patient Reported Outcome Performance Measure.

IV. Endorsement

- Provide the Consensus-Based Entity (CBE) (i.e., Partnership for Quality Measures (PQM)) endorsement status (and CBE ID) and/or other endorsing body. If the measure is only endorsed for paper records, please note endorsement for only the data source being submitted.

This eCQM (CBE ID 4700e) was endorsed with conditions by the Partnership for Quality Measurement (PQM) in March 2025.

V. Summary

- Alignment with CMS Meaningful Measures Initiative or MACRA (if applicable).

This eCQM falls under the following Meaningful Measure 2.0 domains: Closing Gaps of Care; Safety; and Value, Affordability, and Efficiency.

- Relevance to MIPS or other CMS programs.

This eCQM is relevant to MIPS and other Quality Payment Programs where the goals of the programs are to improve the quality of healthcare (processes and patient outcomes) while also reducing the costs of care. This eCQM reports on timely diagnostic resolution after abnormal screening mammograms, which can lead to earlier detection of breast cancer, less invasive treatments, better patient outcomes, and higher quality of life in survivorship.

- Rationale: Use of measure for inclusion in program (specialty society, regional collaborative, other).

Breast cancer is the second most common cause of cancer deaths among women in the United States [1]. A study in early January 2024 predicted that around 42,250 women would die from breast cancer and an estimated 310,720 new cases of invasive breast cancer would be diagnosed that year [1].

Breast cancer survival is dependent upon cancer stage at diagnosis. Approximately 99% of women diagnosed with early-stage breast cancer live for five 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 (Breast Imaging – Reporting and Data System [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 Center for Disease Control and

Prevention 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 [15]. In this cohort, women with no family history of breast cancer had lower follow-up rates, Black and Native American women had lower overall follow-up rates and lower biopsy rates [15]. 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 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.

References:

1. Key Statistics for Breast Cancer. American Cancer Society. Updated January 17, 2024. Accessed October 31, 2024. <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.
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.

- Public reporting (if applicable).

Not applicable, this is a novel eCQM.

- Preferable relevant peer-reviewed journal for publication.

JAMA Internal Medicine.

- Rationale as to how the measure correlates to existing cost measures and improvement activities, as applicable and feasible.

Not applicable.

APPENDIX 1: Individual Clinician Rates

The overall individual clinician eCQM rates for Health System 1 are presented in **Table 15**.

Table 15. eCQM Rates Overall for 99 Individual Clinicians at **Health System 1**

| Facility Group | eCQM Rate of Timely Diagnostic Resolution (95% CI) |
|----------------|--|
| Overall | 91.8 (91.6, 92.0) |
| 1 | 70.4 (59.8, 81.0) |
| 2 | 70.5 (65.1, 75.9) |
| 3 | 80.3 (76.5, 84.2) |
| 4 | 81.1 (76.3, 85.8) |
| 5 | 81.2 (77.8, 84.7) |
| 6 | 82.5 (80.5, 84.5) |
| 7 | 82.5 (78.6, 86.5) |
| 8 | 82.7 (75.7, 89.8) |
| 9 | 83.4 (79.6, 87.2) |
| 10 | 83.9 (74.7, 93.0) |
| 11 | 84.1 (81.7, 86.6) |
| 12 | 85.1 (81.4, 88.7) |
| 13 | 85.8 (81.8, 89.8) |
| 14 | 86.7 (85.1, 88.4) |
| 15 | 86.8 (84.8, 88.7) |
| 16 | 87.2 (83.5, 90.9) |
| 17 | 87.3 (85.0, 89.5) |
| 18 | 87.3 (85.3, 89.3) |
| 19 | 87.4 (85.0, 89.8) |
| 20 | 87.5 (85.3, 89.8) |
| 21 | 87.8 (84.1, 91.5) |
| 22 | 88.0 (86.3, 89.7) |
| 23 | 88.3 (85.7, 90.9) |
| 24 | 88.7 (87.3, 90.0) |
| 25 | 88.8 (85.4, 92.3) |
| 26 | 88.9 (87.3, 90.4) |
| 27 | 89.0 (87.7, 90.3) |
| 28 | 89.2 (87.1, 91.3) |
| 29 | 89.4 (87.9, 91.0) |
| 30 | 89.6 (86.8, 92.3) |
| 31 | 89.7 (87.9, 91.5) |
| 32 | 89.9 (88.4, 91.3) |
| 33 | 90.3 (89.1, 91.5) |
| 34 | 90.8 (87.2, 94.4) |
| 35 | 90.9 (88.3, 93.4) |
| 36 | 90.9 (83.3, 98.5) |
| 37 | 91.0 (89.3, 92.7) |
| 38 | 91.1 (89.4, 92.9) |
| 39 | 91.5 (90.0, 92.9) |
| 40 | 91.8 (86.7, 96.9) |
| 41 | 92.0 (88.7, 95.4) |
| 42 | 92.2 (84.8, 99.5) |
| 43 | 92.3 (90.7, 93.9) |
| 44 | 92.4 (90.8, 94.0) |
| 45 | 92.4 (89.5, 95.3) |
| 46 | 92.4 (88.5, 96.4) |
| 47 | 92.5 (85.3, 99.6) |

| | |
|----|-------------------|
| 48 | 92.5 (90.2, 94.9) |
| 49 | 92.6 (87.4, 97.9) |
| 50 | 92.7 (89.8, 95.7) |
| 51 | 92.9 (91.1, 94.6) |
| 52 | 92.9 (91.3, 94.6) |
| 53 | 93.0 (88.4, 97.7) |
| 54 | 93.2 (91.3, 95.0) |
| 55 | 93.2 (91.1, 95.4) |
| 56 | 93.4 (91.8, 95.0) |
| 57 | 93.4 (92.0, 94.8) |
| 58 | 93.4 (91.9, 95.0) |
| 59 | 93.4 (89.9, 97.0) |
| 60 | 93.6 (92.6, 94.5) |
| 61 | 93.8 (92.9, 94.8) |
| 62 | 93.9 (91.2, 96.6) |
| 63 | 94.0 (93.2, 94.9) |
| 64 | 94.1 (92.5, 95.7) |
| 65 | 94.1 (92.8, 95.4) |
| 66 | 94.2 (89.7, 98.7) |
| 67 | 94.2 (92.9, 95.5) |
| 68 | 94.2 (92.2, 96.3) |
| 69 | 94.3 (90.7, 96.8) |
| 70 | 94.4 (91.5, 97.2) |
| 71 | 94.5 (93.0, 95.9) |
| 72 | 94.8 (93.3, 96.3) |
| 73 | 94.8 (93.3, 96.3) |
| 74 | 95.1 (93.3, 97.0) |
| 75 | 95.2 (94.1, 96.4) |
| 76 | 95.3 (90.8, 99.8) |
| 77 | 95.3 (92.2, 98.5) |
| 78 | 95.5 (90.4, 99.9) |
| 79 | 95.5 (93.8, 97.1) |
| 80 | 95.5 (93.9, 97.0) |
| 81 | 95.6 (94.1, 97.1) |
| 82 | 95.8 (94.9, 96.8) |
| 83 | 95.8 (93.8, 97.9) |
| 84 | 95.9 (94.7, 97.0) |
| 85 | 96.0 (94.1, 97.9) |
| 86 | 96.1 (91.7, 99.9) |
| 87 | 96.2 (95.4, 96.9) |
| 88 | 96.7 (96.0, 97.5) |
| 89 | 96.8 (96.1, 97.5) |
| 90 | 96.9 (95.6, 98.2) |
| 91 | 96.9 (94.6, 99.2) |
| 92 | 97.0 (95.4, 98.6) |
| 93 | 97.1 (94.6, 99.6) |
| 94 | 97.3 (94.2, 99.9) |
| 95 | 97.3 (95.5, 99.2) |
| 96 | 97.7 (95.2, 99.9) |
| 97 | 98.2 (97.3, 99.0) |
| 98 | 98.3 (95.8, 99.9) |
| 99 | 98.7 (97.2, 99.9) |

APPENDIX 2: 30-day Follow-up Rates and Accountable-entity Level Reliability Testing

Health System 1 had an overall integrated delivery system 30-day follow-up diagnostic resolution rate of 79.6% (95% CI: 79.3, 79.9). Health System 2 had an overall 30-day diagnostic resolution rate of 87.8% (95% CI: 87.2, 88.3) based on data extracted from unstructured BI-RADS fields.

The facility group rates by year for Health System 1 are presented in **Table 16**.

Table 16. eCQM Rates by Year for Six Hospital-affiliated Facility Groups at **Health System 1**

| Measurement Year | eCQM Rate of Timely Diagnostic Resolution (95% CI) | | | | | |
|------------------|--|----------------------|----------------------|----------------------|----------------------|----------------------|
| | Facility Group 1 | Facility Group 2 | Facility Group 3 | Facility Group 4 | Facility Group 5 | Facility Group 6 |
| Overall | 69.2 (67.8, 70.6) | 70.4 (69.8, 71.0) | 72.2 (71.2, 73.2) | 85.6 (85.0, 86.1) | 80.7 (80.0, 81.5) | 89.6 (89.2, 90.1) |
| 2018 | 92.9 (91.0, 94.7) | 84.8 (83.5, 86.2) | 86.4 (84.7, 88.2) | 85.6 (84.3, 86.8) | 90.3 (88.8, 91.7) | 95.5 (94.7, 96.3) |
| 2019 | 90.9 (88.5, 93.2) | 86.3 (85.1, 87.6) | 84.2 (82.3, 86.2) | 76.2 (74.8, 77.6) | 84.6 (83.1, 86.1) | 93.9 (93.0, 94.7) |
| 2020 | 84.5 (81.3, 87.7) | 70.2 (68.5, 71.9) | 73.4 (70.8, 76.0) | 85.1 (83.7, 86.6) | 90.6 (89.0, 92.2) | 96.7 (96.0, 97.5) |
| 2021 | 83.2 (80.3, 86.0) | 78.0 (76.7, 79.2) | 64.6 (62.1, 67.0) | 92.8 (91.8, 93.7) | 76.7 (75.0, 78.4) | 94.6 (93.8, 95.4) |
| 2022 | 53.1 (49.8, 56.3) | 62.6 (61.1, 64.1) | 73.3 (71.0, 75.7) | 91.3 (90.2, 92.4) | 78.4 (76.6, 80.3) | 90.4 (89.3, 91.5) |
| 2023 | 29.4 (26.2, 32.5) | 47.7 (46.1, 49.3) | 49.2 (46.5, 52.0) | 85.0 (83.7, 86.3) | 62.4 (59.7, 65.1) | 70.2 (68.6, 71.8) |

The SNRs are provided in **Table 17**. Overall, the median SNR was 0.999 (95% CI: 0.997, 0.999) for the six hospital-affiliated facility groups at Health System 1. The minimum SNR was 0.997 and the maximum SNR was 0.999. The SNRs were high across all years from 2018 to 2023. The >0.99 median SNRs with narrow 95% confidence intervals, overall and across all years, indicate that a very high proportion of overall variability is explained by the differences between measured entities (i.e., hospital-affiliated facility groups).

Table 17. Signal-to-Noise Ratios (SNR), Overall and by Year for Six Facility Groups in **Health System 1**

| Measurement Year | Median SNR (95% CI) | Minimum SNR | Maximum SNR |
|------------------|-----------------------------|--------------|--------------|
| Overall | 0.999 (0.997, 0.999) | 0.997 | 0.999 |
| 2018 | 0.994 (0.982, 0.999) | 0.978 | 0.995 |
| 2019 | 0.994 (0.979, 0.999) | 0.975 | 0.996 |
| 2020 | 0.999 (0.991, 0.999) | 0.989 | 0.998 |
| 2021 | 0.998 (0.993, 0.999) | 0.991 | 0.999 |
| 2022 | 0.998 (0.994, 0.999) | 0.993 | 0.999 |
| 2023 | 0.997 (0.994, 0.999) | 0.994 | 0.999 |

The overall Spearman's rank correlation coefficient was 1.00 (95% CI: 0.99, 0.99) (**Table 18**). The >0.94 correlations, overall and across all years, indicated a very strong positive correlation between the test and validation samples. However, some confidence intervals by year were wide given that only six facility groups were included in the analysis. Additional facility group data is required to generate narrower confidence intervals in these cases.

Table 18. Spearman's Rank Correlation Coefficients, Overall and by Year for Six Facility Groups in **Health System 1**

| Measurement Year | Test-Validation Correlation | 95% CI |
|------------------|-----------------------------|---------------------|
| Overall | 1.00 | (0.99, 0.99) |
| 2018 | 1.00 | (0.99, 0.99) |

| | | |
|------|------|--------------|
| 2019 | 0.94 | (0.49, 0.99) |
| 2020 | 0.94 | (0.49, 0.99) |
| 2021 | 0.94 | (0.49, 0.99) |
| 2022 | 1.00 | (0.99, 0.99) |
| 2023 | 1.00 | (0.99, 0.99) |

The overall ICC was 0.078 (95% CI: 0.032, 0.340) in the test sample and 0.092 (95% CI: 0.038, 0.382) in the validation sample (**Table 19**). There were no apparent trends over time. The overall ICCs indicated that a low proportion of variation in the facility group level scores was due to facility group level signal variation. Notably, the 95% CIs were very wide given that only six facility groups were included in the analysis. Additional facility group data is required to generate narrower confidence intervals.

Table 19. Interclass Correlation Coefficients (ICC), Overall and by Year for Six Facility Groups in **Health System 1**

| Measurement Year | Test Sample (95% CI) | Validation Sample (95% CI) |
|------------------|-----------------------------|-----------------------------|
| Overall | 0.078 (0.032, 0.340) | 0.092 (0.038, 0.382) |
| 2018 | 0.070 (0.027, 0.344) | 0.084 (0.033, 0.390) |
| 2019 | 0.072 (0.028, 0.348) | 0.093 (0.037, 0.400) |
| 2020 | 0.200 (0.087, 0.614) | 0.206 (0.090, 0.625) |
| 2021 | 0.175 (0.076, 0.569) | 0.201 (0.089, 0.610) |
| 2022 | 0.181 (0.079, 0.577) | 0.212 (0.095, 0.624) |
| 2023 | 0.183 (0.080, 0.580) | 0.210 (0.093, 0.621) |

APPENDIX 3: 45-day Follow-up Rates and Accountable-entity Level Reliability Testing

Health System 1 had an overall integrated delivery system 45-day follow-up diagnostic resolution rate of 88.2% (95% CI: 88.0, 88.4). Health System 2 had an overall 45-day diagnostic resolution rate of 91.6% (95% CI: 91.1, 92.1) based on data extracted from unstructured BI-RADS fields.

The facility group rates by year for Health System 1 are presented in **Table 20**.

Table 20. eCQM Rates by Year for Six Hospital-affiliated Facility Groups at **Health System 1**

| Measurement Year | eCQM Rate of Timely Diagnostic Resolution (95% CI) | | | | | |
|------------------|--|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | <i>Facility Group 1</i> | <i>Facility Group 2</i> | <i>Facility Group 3</i> | <i>Facility Group 4</i> | <i>Facility Group 5</i> | <i>Facility Group 6</i> |
| Overall | 80.6 (79.4, 81.8) | 81.6 (81.1, 82.2) | 86.5 (85.7, 87.2) | 91.6 (91.1, 92.0) | 91.0 (90.5, 91.5) | 93.5 (93.1, 93.9) |
| 2018 | 95.8 (94.3, 97.2) | 90.1 (88.9, 91.2) | 92.4 (91.0, 93.7) | 89.4 (88.3, 90.6) | 94.3 (93.2, 95.5) | 96.8 (96.1, 97.5) |
| 2019 | 94.4 (92.6, 96.3) | 91.6 (90.5, 92.6) | 91.5 (90.0, 93.0) | 88.3 (87.3, 89.4) | 91.4 (90.2, 92.6) | 5.8 (95.0, 96.5) |
| 2020 | 90.6 (88.1, 93.2) | 83.7 (82.3, 85.0) | 88.0 (86.1, 89.9) | 91.1 (89.9, 92.2) | 94.5 (93.3, 95.8) | 97.4 (96.7, 98.1) |
| 2021 | 91.3 (89.1, 93.4) | 87.7 (86.1, 88.1) | 84.9 (83.1, 86.7) | 95.5 (94.8, 96.3) | 90.4 (89.2, 91.6) | 96.4 (95.8, 97.0) |
| 2022 | 77.8 (75.0, 80.5) | 77.9 (76.6, 79.2) | 87.1 (85.3, 88.9) | 94.0 (93.1, 95.0) | 91.0 (89.7, 92.2) | 93.7 (92.9, 94.6) |
| 2023 | 45.3 (41.8, 48.7) | 64.3 (62.7, 65.8) | 73.9 (71.5, 76.3) | 92.0 (91.0, 93.0) | 83.4 (81.3, 85.4) | 82.7 (81.4, 84.0) |

The SNRs are provided in **Table 21**. Overall, the median SNR was 0.999 (95% CI: 0.996, 0.999) for the six hospital-affiliated facility groups at Health System 1. The minimum SNR was 0.995 and the maximum SNR was 0.999. The SNRs were high across all years from 2018 to 2023. The >0.98 median SNRs with narrow

95% confidence intervals, overall and across all years, indicate that a very high proportion of overall variability is explained by the differences between measured entities (i.e., hospital-affiliated facility groups).

Table 21. Signal-to-Noise Ratios (SNR), Overall and by Year for Six Facility Groups in **Health System 1**

| Measurement Year | Median SNR (95% CI) | Minimum SNR | Maximum SNR |
|------------------|-----------------------------|--------------|--------------|
| Overall | 0.999 (0.996, 0.999) | 0.995 | 0.999 |
| 2018 | 0.993 (0.978, 0.999) | 0.973 | 0.994 |
| 2019 | 0.987 (0.952, 0.999) | 0.941 | 0.990 |
| 2020 | 0.995 (0.982, 0.999) | 0.978 | 0.997 |
| 2021 | 0.995 (0.983, 0.999) | 0.979 | 0.997 |
| 2022 | 0.996 (0.989, 0.999) | 0.987 | 0.998 |
| 2023 | 0.997 (0.992, 0.999) | 0.991 | 0.999 |

The overall Spearman's rank correlation coefficient was 1.00 (95% CI: 0.99, 0.99) (**Table 22**). The Spearman's rank correlation coefficient for 2023, which is most reflective of current performance, was 0.94 (95% CI: 0.49, 0.99). The overall Spearman's rank correlation coefficient indicated a very strong positive correlation between the test and validation samples. However, confidence intervals by year were wide given that only six facility groups were included in the analysis. Additional facility group data is required to generate narrower confidence intervals.

Table 22. Spearman's Rank Correlation Coefficients, Overall and by Year for Six Facility Groups in **Health System 1**

| Measurement Year | Test-Validation Correlation | 95% CI |
|------------------|-----------------------------|---------------------|
| Overall | 1.00 | (0.99, 0.99) |
| 2018 | 1.00 | (0.99, 0.99) |
| 2019 | 0.77 | (-0.18, 0.97) |
| 2020 | 0.94 | (0.49, 0.99) |
| 2021 | 0.87 | (0.18, 0.99) |
| 2022 | 1.00 | (0.99, 0.99) |
| 2023 | 0.94 | (0.49, 0.99) |

The overall ICC was 0.068 (95% CI: 0.027, 0.308) in the test sample and 0.079 (95% CI: 0.032, 0.345) in the validation sample (**Table 23**). There were no apparent trends over time. The overall ICCs indicated that a low proportion of variation in the facility group level scores was due to facility group level signal variation. Notably, the 95% CIs were very wide given that only six facility groups were included in the analysis. Additional facility group data is required to generate narrower confidence intervals.

Table 23. Interclass Correlation Coefficients (ICC), Overall and by Year for Six Facility Groups in **Health System 1**

| Measurement Year | Test Sample (95% CI) | Validation Sample (95% CI) |
|------------------|-----------------------------|-----------------------------|
| Overall | 0.068 (0.027, 0.308) | 0.079 (0.032, 0.345) |
| 2018 | 0.066 (0.025, 0.349) | 0.075 (0.029, 0.375) |
| 2019 | 0.044 (0.016, 0.286) | 0.040 (0.015, 0.249) |
| 2020 | 0.136 (0.056, 0.516) | 0.114 (0.046, 0.478) |
| 2021 | 0.104 (0.042, 0.432) | 0.108 (0.044, 0.444) |
| 2022 | 0.098 (0.040, 0.413) | 0.143 (0.060, 0.514) |
| 2023 | 0.195 (0.085, 0.599) | 0.207 (0.092, 0.618) |