

CBE ID

4220

Title

Breast Cancer Screening Recall Rates

Project

Initial Recognition and Management

Endorsement Status

Not Endorsed

E&M Committee Rationale/Justification

Due to No Consensus.

Is Under Review

No

Previous Endorsement Cycle

Fall 2023

Steward

Centers for Medicare & Medicaid Services

1.0 New or Maintenance

New

1.3 Electronic Clinical Quality Measure (eCQM)

No

1.6 Measure Description

The *Breast Cancer Screening Recall Rates* measure calculates the 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.

1.7 Composite Measure

No

1.7 Measure Type

Process

1.8 Level of Analysis

Facility

1.9 Care Setting

Hospital: Outpatient

1.10 Measure Rationale

The Breast Cancer Screening Recall Rates measure calculates the 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 on the same day or within 45 days.

1.11 Measure Webpage

https://qualitynet.cms.gov/files/61e6c24c5277d800221ca92f?filename=OP-39_Ad-Hoc...

1.13 Data Dictionary

Attached

1.13a Attach Data Dictionary

[4220e Value Sets.zip](#)

1.14 Numerator

Medicare beneficiaries who had a diagnostic mammography study, DBT, ultrasound, or MRI of the breast following a screening mammography or DBT study on the same day or within 45 days of the screening study in any location.

1.14a Numerator Details

CBE #4220e calculates the percentage of mammography and 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. The measure's denominator contains any Medicare beneficiary who underwent a screening mammography or DBT study at a facility subject to OPPS regulation during the measurement period. From these beneficiaries, the numerator contains beneficiaries who had a diagnostic mammography study, DBT, ultrasound, or MRI of the breast following a screening mammography or DBT study within 45 days.

The Current Procedural Terminology (CPT) and Healthcare Common Procedure Coding System (HCPCS) codes used to identify beneficiaries with a diagnostic mammography study, DBT, ultrasound or MRI can be found in the submitted Excel file.

1.15 Denominator

Medicare beneficiaries who underwent a screening mammography or DBT study at a facility reimbursed through the Outpatient Prospective Payment System (OPPS).

1.15a Denominator Details

The CBE #4220e denominator contains any Medicare beneficiary who underwent a screening mammography or screening DBT's performed at a facility subject to Outpatient Prospective Payment System (OPPS) regulation during the measurement period. The CPT and HCPCS codes used to identify beneficiaries who underwent a screening mammography or screening DBT can be found in the submitted Excel file.

1.15b Denominator Exclusions

None.

1.15c Denominator Exclusions Details

None.

1.16 Type of Score

Rate/proportion

1.17 Measure Score Interpretation

Better performance = Score within a defined interval

1.18 Calculation of Measure Score

Please see attached measure score calculation diagram within the attachment under the 'Measure Score Calculation Diagram' question.

1.18a Attach measure score calculation diagram

[4220e Measure Score Diagram.pdf](#)

1.19 Measure Stratification Details

Not applicable—CBE #4220e is not stratified.

1.20 Types of Data Sources

Claims Data

1.25 Data Source Details

CBE #4220e is calculated using data from final claims that facilities submit for Medicare beneficiaries enrolled in fee-for-service (FFS) Medicare. The data are calculated only for facilities paid through the OPPS for mammography and DBT screening in the hospital outpatient setting. Data are pulled from the hospital outpatient and carrier files to identify eligible cases for inclusion in the initial patient population and numerator (e.g., a mammography follow-up study can occur in any location and be included in the measure's numerator). Due to claims adjudication, there is a

lag between when an imaging study is performed and when it is reported on the public reporting website.

1.26 Minimum Sample Size

CBE #4220e uses a relative precision model to determine the minimum necessary number of cases. Similar approaches are used for three other Outpatient Imaging Efficiency measures. For this precision model, calculating minimum case count is determined by acceptable levels of precision, the level of confidence necessary for each measure, and the minimum case count required to meet precision and confidence. Precision depends on the facility's observed performance rate. In general, stricter levels of precision are necessary for scores that are closer to the tail ends of the possible range of the measure score (i.e., 0.05 or 0.95), whereas scores towards the middle of the possible range (e.g., 0.50) do not require as strict a level of precision. The level of significance is 0.10. Thus, the minimum case counts (see *Table 1* within the attachment under the 'Logic Model' question) ensure 90 percent confidence that the observed score reflects the true score of the minimum case counts that would be necessary to publicly report CBE #4220e. Facilities would need at least 31 cases to qualify for public reporting; this number can vary from 31 to 67, depending on a facility's performance rate.

2.1 Attach Logic Model

[4220e Logic Model and Tables.pdf](#)

2.2 Evidence of Measure Importance

From the perspective of both clinical quality and efficiency, there are potentially negative consequences if the mammography and DBT recall rate is either too high or too low. A high cumulative dose of low-energy radiation can be a consequence of too many false-positive mammography and DBT follow-up studies. Radiation received from mammography or DBT may induce more cancers in younger people or those carrying deleterious gene mutations, such as BRCA-1 and BRCA-2 (Berrington de Gonzalez et al., 2009).

Societies and guidelines provide inconsistent suggestions on the appropriate recall rates to establish for breast cancer screening. The ACR recommends a target recall rate for mammography screening between 5 percent and 12 percent (American College of Radiology 2013); European research, via the International Agency for Research on Cancer, sets a target recall rate of 5 percent..

References:

Berrington de Gonzalez, A., Berg, C., Visvanathan, K., & Robson, M. (2009). Estimated risk of radiation-induced breast cancer from mammographic screening for young BRCA mutation carriers. *JNCI: Journal of the National Cancer Institute*, 101(3), 205-209.

<https://doi.org/10.1093/jnci/djn440>

D'Orsi, C. J., Sickles, E. A., Mendelson, E. B., Morris EA, et al. (2013). ACR BI-RADS® atlas, breast imaging reporting and data system. Reston, VA: American College of Radiology.

2.3 Anticipated Impact

This measure will guide breast cancer screening decision making in hospital outpatient departments as there are potentially negative consequences if the mammography and DBT recall rate is either too high or too low.

The measure potentially will reduce radiation received from mammography or DBT that may induce more cancers in younger people or those carrying deleterious gene mutations, as well as decreasing unnecessary imaging and biopsies. Alternatively, underuse of follow up for screening mammography or DBT may result in missed cases of cancer.

CMS calculates performance for its Outpatient Imaging Efficiency measures using data from final claims that facilities submit for Medicare beneficiaries enrolled in FFS Medicare. The data are calculated only for facilities paid through the OPSS for mammography and DBT screening studies in the hospital outpatient setting. Data from the hospital outpatient and carrier files are used to determine beneficiary inclusion (e.g., a mammography follow-up study can occur in any location and be included in the measure's numerator).

Results reported are for the public reporting period based on data collected from July 1, 2021, through June 30, 2022 (referred to as *2023 public reporting* or *PR 2023*). In PR 2023, 3,652 facilities had at least 1 eligible case in the measure denominator. A total of 3,391 facilities met the minimum case count requirement, making them eligible for public reporting.

The analysis of the performance gap is presented in *Table 2* and *Table 3*, within the attachment under the 'Logic Model' question. *Table 2* presents the distribution of performance scores and denominator counts for facilities meeting MCC and for all facilities with at least one case in the denominator. *Table 3* presents measure performance scores by patient biological sex, racial or ethnic identity, age group, and dual eligibility status, including chi-square values and probabilities used to assess whether differences in performance are statistically significant. For these analyses, only cases from facilities meeting minimum case count requirements for public reporting were used.

Table 2 shows the mean measure performance for facilities meeting MCC (8.5 percent, standard deviation [S.D.], 6.7 percent) falls within the targeted recall rate range of 5 percent to 12 percent; however, analysis of performance across deciles demonstrates variability across facilities during the measurement period, with more than 30 percent (33.4) of facilities having scores outside of the targeted recall rate range. Scores for all eligible facilities (i.e., those with at least one case in the denominator) add an additional 261 facilities and 7,475 patients; these facilities display a similar distribution with slightly higher mean performance (8.9 percent, S.D., 8.7 percent).

Performance by patient characteristics displayed in *Table 3* show statistically significant differences in performance by biological sex, racial or ethnic identity, age band, and dual eligibility status. Care should be taken in interpretation of these results as some categories make up a small percentage of the total for each characteristic. For example, only 0.01 percent of patients in the measure sample are male (as would be expected, given the clinical scope of the measure), although the chi-square probability (<0.0001) indicates the difference in performance (24.3 percent for males, 9.2 percent for females) is significant. Racial identity also provides a

similar chi-square probability (<0.0001), with white patients making up the majority of cases (86.4 percent of the total initial patient population, with a performance rate of 9.2 percent) followed by Black patients (7.6 percent of the initial patient population, with a performance rate of 8.5 percent). The next largest category is unknown race, comprising 2.1 percent of the initial patient population, with performance at 10.7 percent. While comprising a small percentage of the initial patient population, performance scores for patients of other race (9.8 percent), Asian or Pacific Islander (10.0 percent), and American Indian or Alaska Native (6.7 percent) show significant variation between race categories. Similarly, patients of Hispanic or Latino (9.4 percent) ethnicity also vary substantially from non-Hispanic or non-Latino populations.

Age band categories show consistent trends of lower scores as age increases, ranging from 17.6 percent for patients aged 18 to 34, to 8.2 percent for those age 85 or older. Younger patients make up a small percentage of the overall testing population, with the categories including those aged 18 to 54 comprising about 2.5 percent of the initial patient population. Those aged 55 to 64 are 4.9 percent of the initial patient population, with performance score of 9.3 percent. Patients over the age of 65 make up 92.6 percent of the initial patient population, with scores ranging from 9.3 percent (for ages 65 to 74) to 8.2 percent (for patients who are 85 or older).

Finally, performance by dual eligibility was examined, with 92.6 percent of the initial patient population having only Medicare FFS coverage, and the remaining 7.4 percent enrolled in both Medicare FFS and Medicaid (dually eligible). The difference in performance was slight—9.2 percent for Medicare only versus 9.3 percent for dual eligible—but significant at the 0.05 level ($p=0.0161$).

2.5 Health Care Quality Landscape

While other measures evaluate rates of breast cancer imaging, CBE #4220e monitors rate of recall following screening imaging. This measure provides valuable information for facilities, clinicians, administrators, policy makers, patients, researchers, and others to identify facilities recalling the appropriate number of patients for follow-up screening each year.

2.6 Meaningfulness to Target Population

Among the target population, additional imaging, and biopsies after a screening mammography or DBT can result in over-diagnosis among patients who do not have breast cancer, increasing their anxiety and distress. Alternatively, inappropriately low recall rates may lead to delayed diagnoses or undetected cases of breast cancer (Nelson et al., 2019). Inclusion of DBT when evaluating follow-up care may improve recall rates and positive prediction values compared to metrics that focus on mammography alone (Aujero et al., 2017; Bian et al., 2016; Chong et al., 2019; Conant et al., 2016; Pozzi et al., 2016; and Skaane 2017).

References:

Aujero, M., Gavenonis, S., Benjamin, R., Zhang, Z., & Holt, J. (2017). Clinical performance of synthesized two-dimensional mammography combined with tomosynthesis in a large screening population. *Radiology*, 283(1), 70-76. <https://doi.org/10.1148/radiol.2017162674>

- Berrington de Gonzalez, A., Berg, C., Visvanathan, K., & Robson, M. (2009). Estimated risk of radiation-induced breast cancer from mammographic screening for young BRCA mutation carriers. *JNCI: Journal of the National Cancer Institute*, 101(3), 205-209. <https://doi.org/10.1093/jnci/djn440>
- Bian, T., Lin, Q., Cui, C., Li, L., Qi, C., Fei, J., & Su, X. (2016). Digital breast tomosynthesis: A new diagnostic method for mass-like lesions in dense breasts. *The Breast Journal*, 22(5), 535-540. <https://doi.org/10.1111/tbj.12622>
- Chong, A., Weinstein, S., McDonald, E., & Conant, E. (2019). Digital breast tomosynthesis: Concepts and clinical practice. *Radiology*, 292(1), 1-14. <https://doi.org/10.1148/radiol.2019180760>
- Conant, E., Beaber, E., Sprague, B., Herschorn, S., Weaver, D., Onega, T., Tosteson, A., McCarthy, A., Poplack, S., Haas, J., Armstrong, K., Schnall, M., & Barlow, W. (2016). Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: A cohort study within the PROSPR Consortium. *Breast Cancer Research and Treatment*, 156(1), 109-116. <https://doi.org/10.1007/s10549-016-3695-1>
- Pozzi, A., Corte, A., Lakis, M., Jeong, H. (2016). Digital breast tomosynthesis in addition to conventional 2D-mammography reduces recall rates and is cost effective. *Asian Pacific Journal of Cancer Prevention*, 17(7), 3521-3526. Retrieved January 11, 2023, from <https://pubmed.ncbi.nlm.nih.gov/27510003>
- Skaane, P. (2016). Breast cancer screening with Digital Breast Tomosynthesis. *Digital Breast Tomosynthesis*, 11-28. https://doi.org/10.1007/978-3-319-28631-0_2

3.1 Contributions Towards Closing Care Gaps

As shown above in *Table 3*, some potential social risk factors were examined to identify performance gaps. These factors include biological sex, racial or ethnic identity, age band, and dual eligibility status. Statistically significant differences in performance have been identified which demonstrate an opportunity for improving health equity based on these risk factors.

4.1 Feasibility Assessment

CBE #4220e was assessed via qualitative survey of a multi-stakeholder group of 32 individuals. The measure developer previously seated a technical expert panel, which contained 12 individuals with extensive experience in clinical care (7 physicians), healthcare administration (3 payers, purchasers, or hospital administration staff), and patients (2 patients who act in an advocacy role). To supplement the information gathered from the technical expert panel, the measure developer also reached out to the American College of Radiology, who provided contact information for clinicians, healthcare administration staff, patients, and caregivers. In total, 25 physicians of various specialties, 5 healthcare administration or management staff, 1 patient, and 2 caregivers responded to the survey (which contained questions about measure face validity, feasibility, and usability). Results from this survey are presented throughout the full measure submission form.

For the question related to feasibility for CBE #4220e, the results indicate that 75 percent of the

respondents agree that the measure does not place an undue burden on hospitals to collect the data. For the individuals that responded either *Disagree* or *Strongly Disagree*, one stated that burden would depend on who is reporting the measure and how it is reported (additional information on measure use appears in the *Use* section below); another person felt the measure would be difficult to track without specific Current Procedural Terminology (CPT) codes to identify diagnostic studies that count as follow-up care in the measure's numerator (which has been resolved); a third respondent felt that burden would be high if exclusion of high-risk individuals was added to the technical specifications (which did not happen); finally, a fourth respondent felt that the measure would have significant burden, but did not explain why.

Results from the qualitative survey related to measure feasibility for CBE #4220e appear in *Table 4*, within the attachment under the 'Measure Logic' question.

4.3 Feasibility Informed Final Measure

No changes were made to the final measure specifications in response to the feasibility assessment. There was high agreement that the measure does not place an undue burden on hospitals to collect the data.

4.4 Proprietary Information

Not a proprietary measure and no proprietary components

4.4a Fees, Licensing, or Other Requirements

There are no fees, licensing, or other requirements to use any aspect of this measure as specified.

5.1.1 Data Used for Testing

As described above, CMS calculates its Outpatient Imaging Efficiency measures using data from final claims that facilities submit for Medicare beneficiaries enrolled in FFS Medicare. The data are calculated only for facilities paid through the OPDS for mammography and DBT screening in the hospital outpatient setting. Data from the hospital outpatient and carrier files are used to determine beneficiary inclusion (e.g., a mammography follow-up study can occur in any location and be included in the measure's numerator).

All reported testing results are for the public reporting period based on data collected from July 1, 2021, through June 30, 2022 (PR 2023). In PR 2023, 3,652 facilities had at least 1 eligible case in the measure denominator. A total of 3,391 facilities met the minimum case count requirement, making them eligible for public reporting.

5.1.2 Differences in Data

The same data were used for all aspects of testing.

5.1.3 Characteristics of Measured Entities

A total of 3,652 facilities were included in the testing population, with 3,315,335 imaging studies included in the measure's denominator. *Table 2* above shows the distribution of performance scores and denominator counts for all facilities as well as for the subset (3,391) of facilities meeting MCC requirements. These include all facilities for which relevant Medicare claims data were available; no sampling strategy was employed.

Distribution for location (i.e., urban versus rural), bed size, teaching status, and ownership status of facilities meeting MCC requirements are shown in *Table 5*, within the attachment under the 'Logic Model' question. The majority of facilities were urban (59.6 percent), non-teaching (83.5 percent), and non-profit (65.8 percent). Distribution by bed size shows a plurality of facilities to be small (0-50 count bed size, 32.5 percent) with substantive proportions at each subsequent bed size category.

5.1.4 Characteristics of Units of the Eligible Population

Table 3, within the attachment under the 'Logic Model' question, displays the distribution of cases included in the denominator from facilities meeting MCC requirements by patient characteristic, including biological sex, racial or ethnic identification, age band, and dual eligibility status.

5.2.1 Level(s) of Reliability Testing Conducted

Accountable entity level (i.e., measure score) (e.g., signal-to-noise analysis)

5.2.2 Method(s) of Reliability Testing

Reliability was calculated in accordance with the methods described in *The Reliability of Provider Profiling: A Tutorial* (2009). This approach calculates the ability of the measure to distinguish between the performances of different facilities. Specifically, the testing calculated the signal-to-noise ratio for each facility meeting minimum case count, with higher scores indicating greater reliability. The reliability score is estimated using a beta-binomial model and is a function of the facility's sample size and score on the measure, as well as the variance across facilities.

References:

Adams, J. (2009). *The Reliability of Provider Profiling: A Tutorial*. <https://doi.org/10.7249/tr653>

5.2.3 Reliability Testing Results

See next section.

5.2.4 Interpretation of Reliability Results

As shown above, reliability scores for CBE #4220e ranged from 0.41 to 1.00, with a median reliability score of 0.95. This median score is indicative of very strong measure reliability and

suggests that this measure is able to identify true differences in performance between individual facilities.

Table 2. Accountable Entity Level Reliability Testing Results by Denominator, Target Population Size

Accountable Entity-Level Reliability Testing Results													
	Overall	Minimum	Decile_1	Decile_2	Decile_3	Decile_4	Decile_5	Decile_6	Decile_7	Decile_8	Decile_9	Decile_10	Maximum
Reliability	0.92	0.41	0.81	0.88	0.91	0.94	0.95	0.97	0.98	0.98	0.99	1.00	1.00
Mean Performance Score	N/A	N/A	339	339	339	341	337	340	339	339	338	340	340
N of Entities	N/A	N/A	41,335	65,887	96,318	124,514	179,137	243,097	319,001	434,845	620,705	1,183,001	1,183,001

5.3.1 Level(s) of Validity Testing Conducted

Accountable entity level (i.e., measure score) (e.g., criterion validity)

5.3.3 Method(s) of Validity Testing

Feedback received from external stakeholders during a listening session about CBE #4220e indicate that a diverse group of stakeholders support its validity. Stakeholders were in agreement that screening mammography and DBT are appropriate imaging modalities that should be used to capture the initial patient population of the measure; some stakeholders recommended the measure consider the addition of other types of screening modalities, such as MRI and ultrasound, to the specifications. Stakeholders also reached consensus that guidance for the measure should include a target screening recall rate of 5 percent to 12 percent, in alignment with American College of Radiology guidelines, noting that the addition of DBT may shift that range downward (because DBT provides more precise imaging). Lastly, stakeholders suggested that facility characteristics could impact recall rates (e.g., underserved areas may have higher recall rates because patients in those areas have limited engagement with the healthcare system and tend to experience fragmented care). Differences in facility characteristics could be corrected using an administrative cancer detection rate.

In addition to the public listening session, the face validity of the measure was systematically assessed via qualitative survey of a multi-stakeholder group of 32 individuals, including one patient/patient advocate (the composition of which is described in the *Feasibility* section, above).

5.3.4 Validity Testing Results

The results shown above in *Table 7*, within the attachment under the 'Logic Model' question, indicate that 75 percent of the respondents support the measure's intent—to assess recall rates to determine appropriate diagnostic imaging for breast cancer detection. Furthermore, 71 percent of the respondents strongly agreed or agreed that the measure addresses quality of care (*Table 8*, within the attachment under the 'Logic Model' question).

For individuals who responded *Disagree* or *Strongly Disagree* to the question summarized in *Table 7*, suggestions were made to update the measure name to reference *recall* in lieu of *follow up* (the change was made) and encouragement to use BIRADS assessment values instead of documentation from CPT codes (which is not feasible for measures calculated using administrative

claims data).

For responses to the question summarized in *Table 8*, one person noted that “there are many additional factors to consider” (without sharing additional context); four individuals suggested removal of magnetic resonance as a follow-up imaging modality (the suggestion was not accepted); and three people encouraged removal of information about breast density (which was addressed).

5.3.5 Interpretation of Validity Results

Please see *Table 7* and *Table 8*, within the attachment under the 'Logic Model' question.

5.3.2 Type of Accountable Entity Level Validity Testing Conducted (derived)

Systematic assessment of face validity of the measure’s performance score as an indicator of quality or resource use

5.4.1 Methods Used to Address Risk Factors

No risk adjustment or stratification

5.4.1b Rationale For No Adjustment or Stratification

CBE #4220e is a process measure for which the measure steward provides no risk adjustment or risk stratification. It was determined that risk adjustment and risk stratification were not appropriate for the measure based on the measure evidence base and the measure construct. Stakeholder feedback received during the listening session for the measure suggest facility characteristics could potentially impact measure scores because patients in underserved areas who have limited engagement with the healthcare system may have higher rates of recall. Additionally, the prevalence of cancer in underserved areas is typically higher because their population tends to have limited to no primary care prevention. As a process-of-care measure, the decision to image a patient should not be influenced by sociodemographic status factors; rather, adjustment would risk masking such important inequities in care delivery. Variation across populations is reflective of differences in the quality of care provided to the disparate population included in the measure’s denominator. The measure steward will continue to assess the need for risk adjustment throughout the measure’s lifespan.

6.1.2 Current or Planned Use(s)

Public Reporting, Quality Improvement (Internal to the specific organization)

6.2.1 Actions of Measured Entities to Improve Performance

Usability of the measure was assessed via qualitative survey of a multi-stakeholder group of 32 individuals, including one patient/patient advocate. The results indicate that 77.4 percent of the respondents agree the measure can be used by hospital outpatient departments to guide decision-making and improve healthcare quality and health outcomes. One responder suggested that the measure be used in conjunction with a breast cancer detection rate. Furthermore, 80.6 percent of

the respondents agreed that the measurement of mammography and DBT follow-up rates for breast cancer detection is highly important because reporting the measure results can supply meaningful information to consumers and healthcare providers.

Developer POC email

ccsq-op-measures@acumenllc.com

Measured/accountable entity (reliability and/or validity) methodology and results (if available)

Measured entity (reliability and validity) methodology and results (if available)

Responder for Survey

Patient

The measure developer is different from the measure steward

Yes

Steward Address

Janis Grady
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Windsor Mill, MD 21244
United States

Steward Organization

Centers for Medicare & Medicaid Services

Steward Organization URL

[https://www.cms.gov/medicare/quality/initiatives/hospital-quality-initiative/ h...](https://www.cms.gov/medicare/quality/initiatives/hospital-quality-initiative/h...)

Steward POC email

Janis.Grady@cms.hhs.gov