NATIONAL QUALITY FORUM
Driving measurable health improvements together

Memo

November 17, 2020

To: Consensus Standards Approval Committee (CSAC)

From: Cancer Project Team

Re: Cancer Spring 2020 Measures[a]

CSAC Action Required

The CSAC will review recommendations from the Cancer project at its November 17-18, 2020 meeting and vote on whether to uphold the recommendations from the Committee.

This memo includes a summary of the project, measure recommendations, themes identified and responses to the public and member comments and the results from the National Quality Forum (NQF) member expression of support. The following documents accompany this memo:

1. **Cancer Fall 2019 Draft Report**. The draft report has been updated to reflect the changes made following the Standing Committee's discussion of public and member comments. The complete draft report and supplemental materials are available on the project webpage.
2. **Comment Table**. Staff has identified themes within the comments received. This table lists two comments received during the post-meeting comment period and the NQF/Standing Committee responses.

Background

Cancer is the second most common cause of death in the U.S., exceeded only by heart disease. The National Cancer Institute (NCI) estimated that in 2018, 1.7 million new cases of cancer would be diagnosed in the United States and over 600,000 people will die from the disease. Furthermore, nearly half of all men and one-third of all women in the U.S. will develop cancer during their lifetime. In addition, diagnosis and treatment of cancer has great economic impact on patients, their families, and society. NCI estimated that, in 2010, the costs for cancer care in the U.S. totaled nearly $157 billion and could reach $174 billion in 2020.

Cancer care is complex and provided in multiple settings—hospitals, outpatient clinics, ambulatory infusion centers, radiation oncology treatment centers, radiology departments, palliative and hospice care facilities—and by multiple providers including surgeons, oncologists, nurses, pain management specialists, pharmacists and social workers.

The Cancer Standing Committee oversees NQF's portfolio of Cancer measures that includes measures for hematology, breast cancer, colon cancer, prostate cancer, and other cancer measures. The purpose of this project was to review Cancer measure submitted for endorsement or undergoing maintenance during the spring 2020 cycle.

---

[a] *This memo is funded by the Centers for Medicare and Medicaid Services under contract HHSM-500-2017-00060I Task Order HHSM-500-T0001*

During the Measure Evaluation Web Meeting held on July 10, 2020, the Cancer Standing Committee evaluated one maintenance measure for endorsement consideration. The measure did not pass on validity.

Draft Report

The Cancer Spring 2020 draft report presents the results of the evaluation of one measure considered under the Consensus Development Process (CDP). One measure was not recommended.

The measures were evaluated against the 2019 version of the measure evaluation criteria.

	Maintenance	New	Total
Measures under consideration	1	0	1
Measures not recommended for endorsement	1	0	1
Reasons for not recommending	Importance - 0 Scientific Acceptability - 1 Use - 0 Overall - 0 Competing Measure - 0	Importance - 0 Scientific Acceptability - 0 Use - 0 Overall - 0 Competing Measure – 0	1

CSAC Action Required

Pursuant to the CDP, the CSAC is asked to consider endorsement of one candidate consensus measure.

Measures Not Recommended for Endorsement

(See Appendix B for the Committee's votes and rationale)

- NQF 0508 Diagnostic Imaging Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms (American College of Radiology)

Comments and Their Disposition

NQF received two comments from two member organizations and individuals pertaining to the draft report and to the measures under consideration.

A table of comments submitted during the comment period, with the responses to each comment and the actions taken by the Standing Committee and measure developers, is posted to the Cancer project webpage.

Comments and Committee Responses

Comments about specific measure specifications and rationale were forwarded to the developers, who were invited to respond.

The Standing Committee reviewed all of the submitted comments (general and measure specific) and developer responses.

Measure-Specific Comments

0508 Diagnostic Imaging Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms (American College of Radiology)

The American Geriatrics Society (AGS) wishes to provide comment on NQF 0508 Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms. We note that this measure as written is somewhat confusing. We are unclear if the intent is that screening mammograms should never be labeled as "probably benign" and if in this gray zone should always be further confirmed (by comparing to prior mammograms for example) as benign or not benign. The wording of "inappropriate use of probably benign assessment" was not clear - is there an appropriate use for this category? We also think that the measure should specify this is only for radiologists reading mammograms; we do not see how this would apply to other clinicians.

Measure Steward/Developer Response:

The "probably benign" assessment category (or Breast Imaging Reporting and Data System (BI-RADS)) 3) is reserved for findings with a high probability (≥98%) of being benign and should not be used as a category for indeterminate findings. Inappropriate designation of findings as "probably benign" can result in the unnecessary follow-up of lesions that could be quickly classified or delayed diagnosis and treatment of cancerous lesions. It is of further note that NQF 0508 guidance documents, specifically BIRADS 3, was commented on during the September 14, 2020, NQF Reducing Diagnostic Error Project meeting. The comment specifically focused on the importance of appropriateness measures improving diagnostic accuracy.

During the Standing Committee meeting convened on July 10, 2020, ACR noticed inconsistencies regarding the Standing Committee's interpretation of the measure as it applies to implementation and measurement elements. Particularly, the Standing Committee's recommendation regarding re-specifying the measure to capture "follow-up mammograms vs first-time mammograms only", rather than the current specifications. ACR respectfully disagrees that the measure should be re-specified for "follow-up" mammograms only. It appears there was not a clear understanding as to the distinction between a screening mammogram and a diagnostic mammogram. While screening mammograms are routinely administered at regular intervals to detect breast cancer in asymptomatic patients, diagnostic mammograms are used after abnormal or suspicious results on a screening mammogram or after some signs of breast cancer alert the physician to check the tissue. Regular interval screening mammograms after an initial or baseline exam are not technically considered "follow-up" exams, thus the recommendation to only capture "follow-up" exams in the measure denominator would be excluding patients with baseline screening mammograms for whom it would also not be advantageous to recommend.

The current measure evaluates assessment of findings on annual or bi-annual, regular interval screening mammography exams as to whether abnormal or suspicious findings are followed up efficiently and expeditiously. Because a "probably benign" BI-RADS 3 assessment at screening defers the diagnostic workup by six months, it is strongly recommended that BI-RADS 3 assessments are issued only on a diagnostic mammography exam after an appropriate workup. Thus, to only capture regular interval screening exams in the measure denominator would be excluding patients with baseline or "first-time" screening mammograms for whom it would also be disadvantageous to defer diagnostic workup. This recommendation has been made based on the following studies which indicate the major advantages that full diagnostic imaging evaluation will provide, in addition to identifying both benign and malignant lesions promptly instead of waiting for six months.

1. More prompt identification of truly benign findings (simple cysts, some intramammary lymph nodes, some cases of grouped skin calcifications, etc.). A large-scale Breast Cancer Surveillance Consortium (BSCS) study, involving more than 1 million mammograms, has shown that recall imaging significantly increases the identification of characteristically benign lesions, thus promptly establishing a benign diagnosis, eliminating 6 months of potential anxiety, and obviating short-interval follow-up examination. (Kerlikowske K, Smith-Bindman R, Abraham LA, et al.

2. More prompt identification of some rapidly growing cancers (the same BCSC study also suggested that recall imaging leads to the prompt diagnosis of some aggressively growing cancers by identifying these tumors when they are smaller and more likely to be node-negative, rather than six months later at initial short-interval follow-up examination.)

NQF 0508 involves reporting the percentage of screening (as opposed to diagnostic) mammography examinations that are assessed as BI-RADS category 3, with the stated goal of reducing this to "approaching 0%" in clinical practice, a BI-RADS category 3 assessment rendered from a screening exam, without prompt diagnostic workup, is considered a positive screening exam. The rationale for making BI-RADS category 3 at screening positive is that it recommends additional imaging evaluation prior to routine screening in one year. Use of BI-RADS category 3 assessment at screening is no longer a strategy to reduce recall rate

ACR also plans to determine the necessary data elements to identify disparities that demonstrate a larger performance gap and examine the performance variance among larger and smaller practices (e.g., low volume: 20 patients and high volume: 100 patients).

We acknowledge that NQF 0508 did not meet NQF's must-pass criterion to achieve appropriate empirical validity evidence based on the testing data submitted, which hypothesized that physicians who perform well on NQF 0508 would also perform well on the other related measures. Unfortunately, we did not find a strong correlation for performance between these measures using the construct validity method. However, ACR plans to re-assess the methodology appropriate for establishing validity and reanalyze the data collected for NQF 0508 once the measure is updated appropriately, following potential revisions associated with the previously mentioned Standing-Committee feedback. Such specification updates and validity testing methodology could present a strong justification for this measure's endorsement.

Committee Response:

The Committee reviewed the comments and the developer's response during its deliberations at the post-comment on October 5, 2020. The Committee did not have any dissenting views or opinions related to the developer's response.

Member Expression of Support

Throughout the 16-week continuous public commenting period, NQF members had the opportunity to express their support ('support' or 'do not support') for each measure submitted for endorsement consideration to inform the Committee's recommendations. No NQF members provided their expression of support or non-support. Appendix C details the expression of support.

Appendix A: CSAC Checklist

The table below lists the key considerations to inform the CSAC's review of the measures submitted for endorsement consideration.

Key Consideration	Yes/No	Notes
Were there any process concerns raised during the CDP project? If so, briefly explain.	No	
Did the Standing Committee receive requests for reconsideration? If so, briefly explain.	No	
Did the Standing Committee overturn any of the Scientific Methods Panel's ratings of Scientific Acceptability? If so, state the measure and why the measure was overturned.	No	
If a recommended measure is a related and/or competing measure, was a rationale provided for the Standing Committee's recommendation? If not, briefly explain.	N/A	
Were any measurement gap areas addressed? If so, identify the areas.	No	
Are there additional concerns that require CSAC discussion? If so, briefly explain.	No	

Appendix B: Measures Not Recommended for Endorsement

The table below lists the Committee's vote and rationale for measures not recommended for endorsement.

Legend: H = High; M = Moderate; L = Low; I = Insufficient

Measure	Voting Results	Standing Committee Rationale
0508 Diagnostic Imaging Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms (American College of Radiology)	**Evidence** H-1; M-16; L-0; I-0 **Gap** H-0; M-7; L-10; I-0 **Reliability** H-15; M-1; L-8; I-9 **Validity** H-0; M-4; L-5; I-8 **Feasibility** H-X; M-X; L-X; I-X **Usability and Use** *Use* Pass-X; No Pass-X *Usability* H-X; M-X; L-X; I-X	The developer provided reliability testing; a beta-binomial model measuring the ratio of signal-to-noise was provided showing a reliability statistic of 0.99 for physicians having a minimum of 10 events for the period 2015-2018, suggesting the measure has high reliability. This Committee agreed with this assessment citing it to be reliable. The developer conducted construct validity, calculating Pearson's coefficients; however, a correlation between this measure and two other process measures was not found. The hypothesis was that good performance on this measure likely indicates that physicians who follow guidelines are working within practices that have good systems for tracking patients or do not unnecessarily recall patients. One Committee member noted that perfect correlation isn't always preferred as it could require a stronger rationale for the need for separate measures. The Committee mentioned that face validity data included by the measure developer showed high face validity; most Committee members seemed to agree with this assessment. It also was acknowledged, however, that as a maintenance measure, NQF requires empiric validity testing. After deliberating on these issues, the Committee passed the measure on reliability, but did not pass the measure on validity.

Appendix C: NQF Member Expression of Support Results

No NQF members provided their expression of support.

Appendix D: Details of Measure Evaluation

Rating Scale: H=High; **M**=Moderate; **L**=Low; **I**=Insufficient; **NA**=Not Applicable

Measures Not Recommended

0508 Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

Submission

Description: Percentage of final reports for screening mammograms that are classified as "probably benign"
Numerator Statement: Final reports classified as "probably benign"
Denominator Statement: All final reports for screening mammograms
Exclusions: No Denominator Exclusions or Denominator Exceptions
Adjustment/Stratification:
Level of Analysis: Clinician : Individual
Setting of Care: Outpatient Services
Type of Measure: Process
Data Source: Claims, Registry Data
Measure Steward: American College of Radiology (ACR)

STANDING COMMITTEE MEETING 07/10/2020
1. Importance to Measure and Report: The measure does not meet the Importance criteria
(1a. Evidence, 1b. Performance Gap)
1a. Evidence: **H-1**; **M-16**; **L-0**; **I-0** 1b. Performance Gap: **H-0**; **M-7**; **L-10**; **I-0**
Rationale:

- The developer noted that it provided updated evidence to include the ACR BI-RADS Atlas, which provides guidance on using a "probably benign" category versus other categories.
- The Committee had a specific question about the logic model presented within the evidence document, which states that if an abnormality is not malignant and the radiologist is also not 100% sure that it is benign, an evaluation of a patient's prior mammography exams is required, rather than an additional diagnostic scan. The developer confirmed that prior mammography exams are used to resolve issues related to abnormal diagnostic tests and mentioned the challenges with diagnostic tests among patients with dense tissue is a frequent topic of discussion among radiologists.
- A Committee member noted that it is important to mention that the measure is more appropriate for follow-up mammograms, rather than first time mammograms, due to the logic model rationale for abnormal diagnostic tests: If an abnormal result was detected, having a prior mammogram for comparison would not be possible for patients who were receiving their first mammogram.
- During the discussion of performance gap, the Committee questioned the BI-RADS categories and scales that determine whether the measure was met. The measure description indicates screening, but the scale in the BI-RADS manual allows for the selection of "3", which is based on a diagnostic test, not a screening. The developer clarified that the distinction for using BI-RADS 3 is only based on screening and diagnostic mammograms, rather than the first or follow-up mammograms. An indication of BI-RADS 3 as "benign" should not routinely be used on a screening exam unless there are unusual circumstances and additional diagnostic tests are recommended.
- The Committee also discussed the average performance across all physicians from 2015-2018, which was 0.52%, with the data from more than 100,000 providers with at least 10 patients who received a mammogram. The measure developer clarified the interpretation of the performance rate, which uses inverse terminology and therefore *not* meeting the measure is the correct quality action. For this measure, 97% of the providers were not compliant.
- The Committee commented on the data from the CMS Physician Quality Reporting System provided by the developer, specifically the difference in the number of physicians who were compliant between the 25% and 75% percentile, which is approximately 170,000. Some members noted this difference is important when deciding if the gap is low or moderate.
- The Committee asked about the performance from physicians with low volume or the variance in performance for physicians who have 20 patients or 100 patients, as this should be viewed differently. The developer responded that it does not receive that level of data from CMS, but would try to review

0508 Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

the performance of physicians within the national mammography database and compare it to those physicians who report through the Merit-based Payment System (MIPS) to measure any variance.
- The Committee's discussion of the performance gap continued as it reviewed the guidelines for qualified mammography centers, citing that the 10-mammogram requirement for this measure seems low.
- The developer shared that the mammography quality standards act requires U.S. radiologists to interpret at least 960 mammograms within a two-year period to be certified. Based on this information, the Committee considered whether the gap was in fact smaller than what was currently shown and how that would equate to the actual numbers of patients for which this measure could be applied.
- The Committee agreed overall that this measure highlights what could be potential harm, such as the physical and emotional harm of a delayed diagnosis.
- The Committee's vote on evidence passed; however, the vote on performance gap did not reach consensus.

2. Scientific Acceptability of Measure Properties: <u>The measure does not meet the Scientific Acceptability criteria</u>

(2a. Reliability precise specifications, testing; 2b. Validity testing, threats to validity)

2a. Reliability: **H-15**; **M-1**; **L-0**; **I-0**; 2b. Validity: **H-0**; **M-4**; **L-5**; **I-8**

<u>Rationale</u>:

- The developer provided reliability testing; a beta-binomial model measuring the ratio of signal-to-noise was provided showing a reliability statistic of 0.99 for physicians having a minimum of 10 events for the period 2015-2018, suggesting the measure has high reliability. This Committee agreed with this assessment citing it to be reliable.
- The developer conducted construct validity, calculating Pearson's coefficients; however, a correlation between this measure and two other process measures was not found. The hypothesis was that good performance on this measure likely indicates that physicians who follow guidelines are working within practices that have good systems for tracking patients or do not unnecessarily recall patients.
- One Committee member noted that perfect correlation isn't always preferred as it could require a stronger rationale for the need for separate measures.
- The Committee mentioned that face validity data included by the measure developer showed high face validity; most Committee members seemed to agree with this assessment. It also was acknowledged, however, that as a maintenance measure, NQF requires empiric validity testing.
- After deliberating on these issues, the Committee passed the measure on reliability, but did not pass the measure on validity.

3. Feasibility: <u>The Standing Committee did not vote on this criteria since the measure did not pass Scientific Acceptability</u>

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

4. Usability and Use: <u>The Standing Committee did not vote on this criteria since the measure did not pass Scientific Acceptability</u>

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: N/A

- The Standing Committee did not recommend the measure for continued endorsement because the measure did not pass validity, a must-pass criterion.

6. Public and Member Comment

- The American Geriatrics Society (AGS) expressed concern that the phrasing of the measure is confusing, particularly the use of "probably benign" as an assessment category. Additional comments addressed concerns around when the measure should be used and how the measure considers mammogram screening intervals.
- The American College of Radiology (ACR) provided a comment that conveyed its disagreement with the Standing Committee's recommendation to re-specify the measure for "follow-up" mammograms only.

0508 Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms
Comment also described ACR's intention to determine the necessary data elements that are necessary to identify disparities and re-assess the methodology appropriate for establishing validity.
7. Consensus Standards Approval Committee (CSAC) Endorsement Decision: Yes-X; No-X
8. Appeals

NATIONAL QUALITY FORUM
Driving measurable health improvements together

http://www.qualityforum.org

Cancer
Spring 2020 Review Cycle

CSAC Review and Endorsement

November 17, 2020

This presentation is funded by the Centers for Medicare and Medicaid Services under contract HHSM-500-2017-00060I Task Order HHSM-500-T0001

Standing Committee Recommendations

- One measure reviewed for Spring 2020
  - Measure was not reviewed by the Scientific Methods Panel
- One measure not recommended for endorsement
  - **NQF 0508** Diagnostic Imaging: Inappropriate Use of 'Probably Benign' Assessment Category in Screening Mammograms Measure Name (Maintenance Measure)

Public and Member Comment and Member Expressions of Support

- Two comments received
- No NQF member expressions of support or non-support received

NATIONAL QUALITY FORUM

Questions?

- Project team:
  - Nicole Williams, MPH, Director
  - Matthew Pickering, PharmD, Senior Director
  - Tamara Funk, MPH, Manager
  - Oroma Igwe, MPH, Manager
  - Teja Vemuganti, MPH, Analyst
  - Mike DiVecchia, MBA, PMP, Project Manager
- Project webpage: http://www.qualityforum.org/Cancer.aspx
- Project email address: cancerem@qualityforum.org

THANK YOU.

NATIONAL QUALITY FORUM

http://www.qualityforum.org

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Cancer, Spring 2020 Cycle: CDP Report

DRAFT REPORT FOR CSAC REVIEW
NOVEMBER 17, 2020

This report is funded by the Centers for Medicare and Medicaid Services under contract HHSM-500-2017-00060I Task Order HHSM-500-T0001

http://www.qualityforum.org

Contents

Executive Summary

Cancer remains a significant burden to patients and the U.S. healthcare system. According to the National Cancer Institute (NCI), an estimated 15.7 million people live with cancer in the United States.[1] In 2020 alone, more than 1.8 million new cancer cases are expected to be diagnosed in the United States and more than 600,000 people will die from the disease.[2] Furthermore, the NCI estimates that the costs for cancer care could reach $174 billion in 2020.[3]

Cancer care is complex and provided in multiple settings—hospitals, outpatient clinics, ambulatory infusion centers, radiation oncology treatment centers, radiology departments, palliative and hospice care facilities—and by multiple providers including surgeons, oncologists, nurses, pain management specialists, and social workers. Due to the complexity of cancer, as well as the numerous care settings and providers, there is a need for quality measures that address the value and efficiency of cancer care for patients and their families.

The National Quality Forum's (NQF) portfolio of measures for cancer includes measures addressing cancer screening and appropriate cancer treatment (including surgery, chemotherapy, and radiation therapy).

For this project, the Cancer Standing Committee evaluated one measure undergoing maintenance review against NQF's standard measure evaluation criteria. The Committee did not recommend this measure for endorsement.

The Committee did not recommend the following measure:

- **NQF 0508** Diagnostic Imaging: Inappropriate Use of 'Probably Benign' Assessment Category in Screening Mammograms (American College of Radiology)

A brief summary of the Committee's evaluation of the measure is included in the body of the report; a detailed summary of the Committee's discussion and ratings of the criteria for the measure are in Appendix A.

Introduction

Cancer is the second most common cause of death in the United States, exceeded only by heart disease.[2] NCI estimates that in 2020, 1.8 million new cases of cancer would be diagnosed in the United States and over 600,000 people will die from the disease.[2] Furthermore, nearly 40% of all men and women in the United States will develop cancer during their lifetime.[4] In addition, diagnosis and treatment of cancer has great economic impact on patients, their families, and the U.S. healthcare system. For 2020, NCI estimates that the costs for cancer care totaled could reach $174 billion.[3]

Given these data points, cancer continues to be recognized as a national priority for quality improvement from the U.S. Department of Health and Human Services, the Centers for Medicare & Medicaid Services (CMS), and numerous other healthcare stakeholders, including commercial payers and medical professional societies.[5–7] These organizations are actively engaged in strategies to address quality of care issues in cancer, including the development and use of quality measures.

Quality measurement in cancer is increasing in significance as the U.S. healthcare system continues to shift from volume to value. In oncology specifically, value-based payment models include financial incentives for adhering to clinical practice guidelines, bundled payments, accountable care organizations, patient-centered medical homes, and the new Oncology Care First Model.[8,9] These models of care have created a demand for measures that can address existing critical quality of care gaps and assess patient experience and quality of life across a range of cancers.

NQF Portfolio of Performance Measures for Cancer Conditions

The Cancer Standing Committee (Appendix C) oversees NQF's portfolio of Cancer measures (Appendix B) that includes measures for hematology, breast cancer, colon cancer, prostate cancer, and other cancer measures. This portfolio contains 18 measures: 17 process measures and one outcome and resource use measure (see table below).

Table 1. NQF Cancer Portfolio of Measures

	Process/Structure	Outcome
Breast Cancer	8	0
Colon Cancer	4	0
Prostate Cancer	2	0
Other Cancer Measures	3	1
Total	17	1

Additional measures related to cancer care are assigned to other portfolios. The additional measures address appropriateness of care (Geriatrics and Palliative Care), cancer screening (Prevention and Population Health), screening for pain, pain related to chemotherapy or radiation therapy, and surgical care.

Cancer Measure Evaluation

On July 10, 2020, the Cancer Standing Committee evaluated one measure undergoing maintenance review against NQF's standard measure evaluation criteria.

Table 2. Cancer Measure Evaluation Summary

	Maintenance	New	Total
Measures under consideration	1	0	1
Measures not recommended for endorsement	1	0	1

Comments Received Prior to Committee Evaluation

NQF solicits comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). In addition, NQF solicits comments for a continuous 16-week period during each evaluation cycle via an online tool located on the project webpage. For this evaluation cycle, the commenting period opened on May 11, 2020 and closed on September 14, 2020. Pre-meeting commenting closed on June 19, 2020. As of that date, no comments were submitted (Appendix F).

Comments Received After Committee Evaluation

The continuous 16-week public commenting period with NQF member support closed on September 15, 2020. Following the Committee's evaluation of the measures under consideration, NQF received two comments from two member organizations and individuals pertaining to the draft report and to the measures under consideration. All comments for each measure under consideration have been summarized in Appendix A.

Throughout the 16-week continuous public commenting period, NQF members had the opportunity to express their support ('support' or 'do not support') for each measure submitted for endorsement consideration to inform the Committee's recommendations. No NQF members provided their expression of support or non-support.

Summary of Measure Evaluation

The following brief summary of the measure evaluation highlights the major issues that the Committee considered. Details of the Committee's discussion and ratings of the criteria for each measure are included in Appendix A.

0508 Diagnostic Imaging: Inappropriate Use of 'Probably Benign' Assessment Category in Screening Mammograms (American College of Radiology): Not Recommended

Description: Percentage of final reports for screening mammograms that are classified as "probably benign"; **Measure Type**: Process; **Level of Analysis**: Clinician : Individual; **Setting of Care**: Outpatient Services; **Data Source**: Claims, Registry Data

The Standing Committee did not vote on the recommendation for endorsement because the Committee did not pass the measure on validity—a must-pass criterion.

The Committee began its discussion with evidence, which was updated by the developer to include the ACR B-RADS Atlas, which provides guidance on using a "probably benign" category versus other categories. The Committee also discussed the logic model presented within the evidence document to describe the steps/decision process when implementing this measure. Specifically, the logic model states that if an abnormality is not malignant and the radiologist is also not 100% sure that it is benign, an evaluation of a patient's prior mammography exams is required, rather than an additional diagnostic scan. The developer clarified that the recommendation is to use prior mammography exams to resolve issues. Based on that information, a Committee member noted that it is important to capture that this measure is applicable to follow-up mammograms rather than first-time mammograms. Overall, the Committee agreed with the evidence provided.

At the outset of the discussion on performance gap, NQF shared the preliminary analysis rating of low for this criterion, which indicates the measure is topped out (mean performance reported was 2.93%, lower score is better). NQF noted that such a high-performance rate allowed the Committee to consider this measure for Reserve Status. The purpose of Reserve Status is to retain endorsement of reliable and valid measures that have overall high levels of performance so that performance can be monitored, as necessary, to ensure that performance does not decline. NQF noted that Reserve Status should be applied only to highly credible, reliable, and valid measures that have high levels of performance due to quality improvement actions (e.g., not due to documentation practices only).

During the discussion on performance gap, the Committee reviewed the data presented from the developer from more than 100,000 providers with at least 10 patients who received a mammogram each year between 2015-2018. The average performance rate across all physicians was 0.52%. The measure developer clarified the interpretation of the performance rate, which uses inverse terminology and therefore *not* meeting the measure is the correct quality action. Providers' case volume as it relates to implementation of this measure also was noted by the Committee, as rates of compliance for providers with a low case volume could be very different. The Committee did not reach consensus on performance gap.

The Committee reviewed and discussed the measure's reliability testing; a beta-binomial model measuring the ratio of signal-to-noise was provided showing a reliability statistic of 0.99 for physicians having a minimum of 10 events for the period 2015–2018. This suggested the measure has high reliability. This Committee agreed with this assessment, concluding it is reliable.

During the discussion on validity, NQF noted the preliminary analysis rating was insufficient. NQF stated that the developer conducted construct validity, calculating Pearson's coefficients. NQF noted, however, that the developer was unable to find a correlation of this measure with two other process measures (including an NQF-endorsed measure), having hypothesized that good performance on this measure likely indicates physicians who follow guidelines are working within practices that have good systems for tracking patients or do not unnecessarily recall patients. The Committee agreed this measure has high face validity, but also acknowledged that it is not the preferred validity for maintenance measures since NQF requires empiric validity testing. The Committee did not pass NQF #0508 on validity, and therefore this measure is not recommended for endorsement.

Measures Withdrawn from Consideration

Two measures previously endorsed by NQF were withdrawn during the endorsement evaluation process. Endorsement for these measures will be removed.

Table 3. Measures Withdrawn from Consideration

Measure	Reason for withdrawal
NQF 0225 At least 12 regional lymph nodes are removed and pathologically examined for resected colon cancer	Developer is not in a position to adequately address concerns with the lack of empiric validity testing data at this time.
NQF 0559 Combination chemotherapy or chemo-immunotherapy (if HER2 positive), is recommended or administered within 4 months (120 days) of diagnosis for women under 70 with AJCC T1cN0 or stage IB – III hormone receptor negative breast cancer	Developer is not in a position to adequately address concerns with the measure's lack of empiric validity testing data at this time.

References

1 Howlander N, Noone A, Krapcho M, et al. *SEER Cancer Statistics Review (CSR) 1975-2017*. National Cancer Institute. Bethesda, MD; 2020. SEER Cancer Statistics Review (CSR) 1975-2017.

2 SEER Cancer Statistics Review, 1975-2017. SEER. https://seer.cancer.gov/csr/1975_2017/index.html. Last accessed July 2020.

3 Cancer Prevalence and Cost of Care Projections. https://costprojections.cancer.gov/. Last accessed July 2020.

4 Cancer Facts & Figures 2020 | American Cancer Society. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2020.html. Last accessed July 2020.

5 Blues plans to launch high-performance network nationwide in 2021. FierceHealthcare. https://www.fiercehealthcare.com/payer/blues-plans-to-launch-high-performance-network-nationwide-2021. Last accessed July 2020.

6 Commission on Cancer. American College of Surgeons. https://www.facs.org/quality-programs/cancer/coc. Last accessed July 2020.

7 Cancer Care Initiatives. ASCO. https://www.asco.org/practice-policy/cancer-care-initiatives. Published January 21, 2016. Last accessed July 2020.

8 Deloitte. The evolution of oncology payment models: what can we learn from early experiments? https://www2.deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-lshc-evolution-of-oncology-payment-models.pdf. Last accessed July 2020.

9 Oncology Care First Model: informal request for information. https://innovation.cms.gov/files/x/ocf-informalrfi.pdf. Last accessed July 2020.

Appendix A: Details of Measure Evaluation

Rating Scale: H=High; **M**=Moderate; **L**=Low; **I**=Insufficient; **NA**=Not Applicable

Measures Not Recommended

0508 Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

Submission

Description: Percentage of final reports for screening mammograms that are classified as "probably benign"
Numerator Statement: Final reports classified as "probably benign"
Denominator Statement: All final reports for screening mammograms
Exclusions: No Denominator Exclusions or Denominator Exceptions
Adjustment/Stratification:
Level of Analysis: Clinician : Individual
Setting of Care: Outpatient Services
Type of Measure: Process
Data Source: Claims, Registry Data
Measure Steward: American College of Radiology (ACR)

STANDING COMMITTEE MEETING 07/10/2020

1. Importance to Measure and Report: The measure does not meet the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **H-1**; **M-16**; **L-0**; **I-0** 1b. Performance Gap: **H-0**; **M-7**; **L-10**; **I-0**

Rationale:

- The developer noted that it provided updated evidence to include the ACR BI-RADS Atlas, which provides guidance on using a "probably benign" category versus other categories.
- The Committee had a specific question about the logic model presented within the evidence document, which states that if an abnormality is not malignant and the radiologist is also not 100% sure that it is benign, an evaluation of a patient's prior mammography exams is required, rather than an additional diagnostic scan. The developer confirmed that prior mammography exams are used to resolve issues related to abnormal diagnostic tests and mentioned the challenges with diagnostic tests among patients with dense tissue is a frequent topic of discussion among radiologists.
- A Committee member noted that it is important to mention that the measure is more appropriate for follow-up mammograms, rather than first time-mammograms, due to the logic model rationale for abnormal diagnostic tests: If an abnormal result was detected, having a prior mammogram for comparison would not be possible for patients who were receiving their first mammogram.
- During the discussion of performance gap, the Committee questioned the BI-RADS categories and scales that determine whether the measure was met. The measure description indicates screening, but the scale in the BI-RADS manual allows for the selection of "3", which is based on a diagnostic test, not a screening. The developer clarified that the distinction for using BI-RADS 3 is only based on screening and diagnostic mammograms, rather than the first or follow-up mammograms. An indication of BI-RADS 3 as "benign" should not routinely be used on a screening exam unless there are unusual circumstances and additional diagnostic tests are recommended.
- The Committee also discussed the average performance across all physicians from 2015-2018, which was 0.52%, with the data from more than 100,000 providers with at least 10 patients who received a mammogram. The measure developer clarified the interpretation of the performance rate, which uses inverse terminology and therefore *not* meeting the measure is the correct quality action. For this measure, 97% of the providers were not compliant.
- The Committee commented on the data from the CMS Physician Quality Reporting System provided by the developer, specifically the difference in the number of physicians who were compliant between the

0508 Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

25% and 75% percentile, which is approximately 170,000. Some members noted this difference is important when deciding if the gap is low or moderate.

- The Committee asked about the performance from physicians with low volume or the variance in performance for physicians who have 20 patients or 100 patients, as this should be viewed differently. The developer responded that it does not receive that level of data from CMS, but would try to review the performance of physicians within the national mammography database and compare it to those physicians who report through the Merit-based Payment System (MIPS) to measure any variance.
- The Committee's discussion of the performance gap continued as it reviewed the guidelines for qualified mammography centers, citing that the 10-mammogram requirement for this measure seems low.
- The developer shared that the mammography quality standards act requires U.S. radiologists to interpret at least 960 mammograms within a two-year period to be certified. Based on this information, the Committee considered whether the gap was in fact smaller than what was currently shown and how that would equate to the actual numbers of patients for which this measure could be applied.
- The Committee agreed overall that this measure highlights what could be potential harm, such as the physical and emotional harm of a delayed diagnosis.
- The Committee's vote on evidence passed; however, the vote on performance gap did not reach consensus.

2. Scientific Acceptability of Measure Properties: <u>The measure does not meet the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-15**; **M-1**; **L-0**; **I-0**; 2b. Validity: **H-0**; **M-4**; **L-5**; **I-8**

<u>Rationale</u>:

- The developer provided reliability testing; a beta-binomial model measuring the ratio of signal-to-noise was provided showing a reliability statistic of 0.99 for physicians having a minimum of 10 events for the period 2015-2018, suggesting the measure has high reliability. This Committee agreed with this assessment citing it to be reliable.
- The developer conducted construct validity, calculating Pearson's coefficients; however, a correlation between this measure and two other process measures was not found. The hypothesis was that good performance on this measure likely indicates that physicians who follow guidelines are working within practices that have good systems for tracking patients or do not unnecessarily recall patients.
- One Committee member noted that perfect correlation isn't always preferred as it could require a stronger rationale for the need for separate measures.
- The Committee mentioned that face validity data included by the measure developer showed high face validity; most Committee members seemed to agree with this assessment. It also was acknowledged, however, that as a maintenance measure, NQF requires empiric validity testing.
- After deliberating on these issues, the Committee passed the measure on reliability, but did not pass the measure on validity.

3. Feasibility: H-X; M-X; L-X; I-X <u>The Standing Committee did not vote on this criterion since the measure did not pass Scientific Acceptability</u>

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

4. Usability and Use: <u>The Standing Committee did not vote on this criterion since the measure did not pass Scientific Acceptability</u>

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use**: Pass-X; No Pass-X**; 4b. Usability: **H-X; M-X; L-X; I-X**

5. Related and Competing Measures

0508 Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms
• No related or competing measures noted.
Standing Committee Recommendation for Endorsement: N/A The Standing Committee did not recommend the measure for continued endorsement because the measure did not pass validity, a must-pass criterion.
6. Public and Member Comment • The American Geriatrics Society (AGS) expressed concern that the phrasing of the measure is confusing, particularly the use of "probably benign" as an assessment category. Additional comments addressed concerns around when the measure should be used and how the measure considers mammogram screening intervals. • The ACR provided a comment that conveyed its disagreement with the Standing Committee's recommendation to re-specify the measure for "follow-up" mammograms only. Comment also described ACR's intention to determine the necessary data elements that are necessary to identify disparities and re-assess the methodology appropriate for establishing validity.
8. Consensus Standards Approval Committee (CSAC) Endorsement Decision: Yes-X; No-X
9. Appeals

Appendix B: Cancer Portfolio—Use in Federal Programs[a]

NQF #	Title	Federal Programs: Finalized or Implemented as of July 16, 2020
0219	Post Breast Conservation Surgery Irradiation	None
0220	Adjuvant Hormonal Therapy	None
0223	Adjuvant Chemotherapy is Recommended or Administered Within 4 Months (120 Days) of Diagnosis to Patients Under the Age of 80 with AJCC III (Lymph Node Positive) Colon Cancer	Hospital Compare (Implemented)
0225	At Least 12 Regional Lymph Nodes Are Removed and Pathologically Examined for Resected Colon Cancer	Prospective Payment System-Exempt Cancer Hospital Quality Reporting (Considered)
0383	Oncology: Plan of Care for Pain – Medical Oncology and Radiation Oncology (paired with 0384)	Hospital Compare (Implemented); Prospective Payment System-Exempt Cancer Hospital Quality Reporting (Implemented); MIPS Program (Implemented)
0384	Oncology: Medical and Radiation - Pain Intensity Quantified	Prospective Payment System-Exempt Cancer Hospital Quality Reporting (Implemented); MIPS Program (Implemented), Medicaid Promoting Interoperability Program for Eligible Professionals (Implemented)
0385	Colon Cancer: Chemotherapy for AJCC Stage III Colon Cancer Patients	None
0385e	Colon Cancer: Chemotherapy for AJCC Stage III Colon Cancer Patients	None
0387	Breast Cancer: Hormonal Therapy for Stage I (T1b)-IIIC Estrogen Receptor/Progesterone Receptor (ER/PR) Positive Breast Cancer	None
0387e	Breast Cancer: Hormonal Therapy for Stage I (T1b)-IIIC Estrogen Receptor/Progesterone Receptor (ER/PR) Positive Breast Cancer	None
0389	Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients	MIPS Program (Implemented)
0389e	Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients	MIPS Program (Implemented), Medicaid Promoting Interoperability Program for Eligible Professionals (Implemented)
0390	Prostate Cancer: Combination Androgen Deprivation Therapy for High Risk or Very High Risk Prostate Cancer	MIPS Program (Implemented)
0508	Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms	None
0509	Diagnostic Imaging: Reminder System for Screening Mammograms	None

[a] Per CMS Measures Inventory Tool as of 07/16/2020

NQF #	Title	Federal Programs: Finalized or Implemented as of July 16, 2020
0559	Combination Chemotherapy is Recommended or Administered Within 4 Months (120 Days) of Diagnosis for Women Under 70 with AJCC T1cN0M0, or Stage IB - III Hormone Receptor Negative Breast Cancer	Hospital Compare (Implemented)
1857	HER2 Negative or Undocumented Breast Cancer Patients Spared Treatment with HER2-Targeted Therapies	MIPS Program (Implemented)
1858	Trastuzumab Administered to Patients with AJCC Stage I (T1c) – III and Human Epidermal Growth Factor Receptor 2 (HER2) Positive Breast Cancer Who Receive Adjuvant Chemotherapy	MIPS Program (Implemented)
1859	KRAS Gene Mutation Testing Performed for Patients with Metastatic Colorectal Cancer Who Receive Anti-Epidermal Growth Factor Receptor Monoclonal Antibody Therapy	MIPS Program (Implemented)
1860	Patients with Metastatic Colorectal Cancer and KRAS Gene Mutation Spared Treatment with Anti-Epidermal Growth Factor Receptor Monoclonal Antibodies	MIPS Program (Implemented)
1878	HER2 Testing for Overexpression or Gene Amplification in Patients with Breast Cancer	None
2930	Febrile Neutropenia Risk Assessment Prior to Chemotherapy	None

Appendix C: Cancer Standing Committee and NQF Staff

STANDING COMMITTEE

Karen Fields, MD (Co-Chair)
Moffitt Cancer Center
Tampa, Florida

Shelley Fuld Nasso, MPP (Co-Chair)
CEO, National Coalition for Cancer Survivorship
Washington DC

Afsaneh Barzi, MD, PhD
Associate Professor, USC – Norris Cancer Center
Los Angeles, California

Gregary Bocsi, DO, FCAP
University of Colorado Hospital Clinical Laboratory
Denver, Colorado

Brent Braveman, Ph.D, OTR/L, FAOTA
University of Texas M.D. Anderson Cancer Center
Houston Texas

Steven Chen, MD, MBA, FACS
OasisMD
Duarte, California

Matthew Facktor, MD, FACS
Geisinger Medical Center
Danville, Pennsylvania

Heidi Floyd
Patient Advocate
Washington, District of Columbia

Bradford Hirsch, MD
SIGNALPATH
Raleigh, North Carolina

Jette Hogenmiller, PhD, MN, APRN/ARNP, CDE, NTP, TNCC, CEE
Oncology Nurse Practitioner
Idaho Springs, Colorado

Wenora Johnson
Research Advocate, Fight Colorectal Cancer
Joliet, Illinois

J. Leonard Lichtenfeld, MD, MACP

American Cancer Society
Atlanta, Georgia

Stephen Lovell, MS
Seattle Cancer Care Alliance Patient and Advisory Council
Washington, District of Columbia

Jennifer Malin, MD, PhD
Anthem, Inc.
Thousand Oaks, California

Jodi Maranchie, MD, FACS
University of Pittsburgh
Pittsburgh, Pennsylvania

Denise Morse, MBA
Director of Quality and Value Analytics, City of Hope Cancer Center
Duarte, California

Benjamin Movsas, MD
Henry Ford Health System
Detroit, Michigan

Beverly Reigle, PhD, RN
University of Cincinnati College of Nursing
Cincinnati, Ohio

David J. Sher, MD, MPH
UT Southwestern Medical Center
Dallas, Texas

Danielle Ziernicki, PharmD
Dedham Group
New York, New York

NQF STAFF

Sheri Winsper, RN, MSN, MSHA
Senior Vice President, Quality Measurement

Apryl Clark, MHSA
Acting Vice President, Quality Measurement

Sai Ma, MPA, PhD
Managing Director/Senior Technical Expert, Quality Measurement

Nicole Williams, MPH
Director

Tamara Funk, MPH
Manager

Oroma Igwe, MPH
Manager

Teja Vemuganti, MPH
Analyst

Mike DiVecchia, MBA, PMP
Project Manager

Robin Y. Nishimi, PhD
NQF Consultant

Appendix D: Measure Specifications

No measures were recommended for endorsement during the spring 2020 cycle.

Appendix E: Related and Competing Measures (Narrative)

Comparison of NQF 0508 and NQF #2372

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms
2372: Breast Cancer Screening

Steward

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

American College of Radiology (ACR)

2372: Breast Cancer Screening

National Committee for Quality Assurance

Description

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

Percentage of final reports for screening mammograms that are classified as "probably benign"

2372: Breast Cancer Screening

Percentage of women 50-74 years of age who had a mammogram to screen for breast cancer

Type

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

Process

2372: Breast Cancer Screening

Process

Data Source

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

Claims, Registry Data Not applicable

Available at measure-specific web page URL identified in S.1 No data dictionary

2372: Breast Cancer Screening

Claims, Electronic Health Data This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Provider Organizations via NCQA's online data submission system.

No data collection instrument provided Attachment 2372_Breast_Cancer_Screening_Value_Sets-636594894640541618.xlsx

Level

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

Clinician : Individual

2372: Breast Cancer Screening

Health Plan, Integrated Delivery System

Setting

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

Outpatient Services

2372: Breast Cancer Screening

Outpatient Services

Numerator Statement

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

Final reports classified as "probably benign"

2372: Breast Cancer Screening

Women who received a mammogram to screen for breast cancer.

Numerator Details

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

Numerator Definition:

Probably Benign Classification – Mammography Quality Standards Act (MQSA) assessment category of "probably benign"; Breast Imaging-Reporting and Data System (BI-RADS®) category 3; or Food and Drug Administration (FDA)-approved equivalent assessment category

Numerator Instructions:

INVERSE MEASURE - A lower calculated performance rate for this measure indicates better clinical care or control. The "Performance Not Met" numerator option for this measure is the representation of the better clinical quality or control. Submitting that numerator option will produce a performance rate that trends closer to 0%, as quality increases. For inverse measures, a rate of 100% means all of the denominator eligible patients did not receive the appropriate care or were not in proper control.

A lower percentage, with a definitional target approaching 0%, indicates appropriate assessment of screening mammograms.

The mammogram assessment category (corresponding CPT Category II 33xxF code for assessment categories") to be submitted is the final assessment for the screening mammographic study. If a diagnostic mammographic study follows the screening exam, the assessment category for the screening exam should be submitted with the corresponding CPT Category II code, i.e. 3340F for Mammogram assessment category of "incomplete: need additional imaging evaluation," documented. Of note, the performance tags indicating 'Performance Met' and 'Performance Not

Met' are included to highlight what is being measured and submitted and not to encourage the use and documentation of "probably benign".

Numerator Options:

Performance Met: Mammogram assessment category of "probably benign," documented (3343F)

OR

Performance Not Met: Mammogram assessment category of "incomplete: need additional imaging evaluation," documented (3340F)

OR

Performance Not Met: Mammogram assessment category of "negative,"

Documented (3341F)

OR

Performance Not Met: Mammogram assessment category of "benign,"

Documented (3342F)

OR

Performance Not Met: Mammogram assessment category of "suspicious,"

Documented (3344F)

OR

Performance Not Met: Mammogram assessment category "highly suggestive of malignancy," documented (3345F)

OR

Performance Not Met: Mammogram assessment category of "known biopsy proven malignancy," documented (3350F)

2372: Breast Cancer Screening

One or more mammograms any time on or between October 1 two years prior to the measurement year and December 31 of the measurement year.

Notes:

(1) This measure assesses the use of imaging to detect early breast cancer in women. Because the measure denominator does not remove women at higher risk of breast cancer, all types and methods of mammograms (screening, diagnostic, film, digital or digital breast tomosynthesis) qualify for numerator compliance. MRIs, ultrasounds or biopsies do not count toward the numerator; although they may be indicated for evaluating women at higher risk for breast cancer or for diagnostic purposes, they are performed as an adjunct to mammography and do not themselves count toward the numerator.

(2) The numerator time frame is 27 months. NCQA allows for a 3-month leeway, a method used for other HEDIS measures (as determined on a per-measure basis), in recognition of the logistics of referrals and scheduling and to avoid potential overuse of screening. This time frame was recommended by our expert advisory panels and approved by our Committee on Performance Measurement, which oversees measures used in the HEDIS Health Plan Measures Set.

See attached code value sets.

Denominator Statement

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

All final reports for screening mammograms

2372: Breast Cancer Screening

Women 50-74 years of age.

Denominator Details

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

Denominator Criteria (Eligible Cases):

Diagnosis for screening mammogram (ICD-10-CM): Z12.31

AND

Patient procedure during the performance period (CPT or HCPCS): 77067

2372: Breast Cancer Screening

Women 52-74 years as of the end of the measurement year (December 31).

Note: this denominator statement captures women age 50-74 years; it is structured to account for the look-back period for mammograms.

Exclusions

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

No Denominator Exclusions or Denominator Exceptions

2372: Breast Cancer Screening

This measure excludes women with a history of bilateral mastectomy. The measure also excludes patients who use hospice services or are enrolled in an institutional special needs plan or living long-term in an institution any time during the measurement year.

Exclusion Details

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

None

2372: Breast Cancer Screening

Exclude patients with bilateral mastectomy any time during the member's history through December 31 of the measurement year. Any of the following meet criteria for bilateral mastectomy:

1) Bilateral mastectomy (Bilateral Mastectomy Value Set)

2) Unilateral mastectomy (Unilateral Mastectomy Value Set) with a bilateral modifier (Bilateral Modifier Value Set)

3) Two unilateral mastectomies (Unilateral Mastectomy Value Set) with service dates 14 days or more apart

4) History of bilateral mastectomy (History of Bilateral Mastectomy Value Set)

5) Any combination of codes that indicate a mastectomy on both the left and right side on the same or different dates of service. Left mastectomy includes any of the following: unilateral mastectomy (Unilateral Mastectomy Value Set) with a left-side modifier (Left Modifier Value Set) same claim; or absence of the left breast (Absence of Left Breast Value Set); or left unilateral mastectomy (Unilateral Mastectomy Left Value Set). Right Mastectomy includes any of the following: unilateral mastectomy (Unilateral Mastectomy Value Set) with a right-side modifier (Right Modifier Value Set) same claim; or absence of the right breast (Absence of Right Breast Value Set); or right unilateral mastectomy (Unilateral Mastectomy Right Value Set).

Exclude patients who use hospice services any time during the measurement year (Hospice Value Set).

Exclude patients 65 and older who are enrolled in an institutional SNP or living long-term in an institution at any time during the measurement year.

Risk Adjustment

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

No risk adjustment or risk stratification

No - This measure is not risk-adjusted

2372: Breast Cancer Screening

No risk adjustment or risk stratification

No - This measure is not risk-adjusted

Stratification

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

We encourage the results of this measure to be stratified by race, ethnicity, sex, and payer.

2372: Breast Cancer Screening

N/A

Type Score

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

Rate/proportion better quality = lower score

2372: Breast Cancer Screening

Rate/proportion better quality = higher score

Algorithm

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

To calculate performance rates:

1) Find the patients who meet the initial patient population (ie, the general group of patients that the performance measure is designed to address).

2) From the patients within the initial patient population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial patient population and denominator are identical.

3) From the patients within the denominator, find the patients who qualify for the Numerator (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator

If the patient does not meet the numerator, this case represents a quality failure. To calculate performance rates:

1) Find the patients who meet the initial patient population (ie, the general group of patients that the performance measure is designed to address).

2) From the patients within the initial patient population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial patient population and denominator are identical.

3) From the patients within the denominator, find the patients who qualify for the Numerator (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator

If the patient does not meet the numerator, this case represents a quality failure.

2372: Breast Cancer Screening

Step 1. Determine the eligible population: identify women 52-74 years of age by the end of the measurement year.

Step 2. Search for an exclusion: history of bilateral mastectomy; or use of hospice services during the measurement year; or patients 65 and older who are enrolled in an institutional SNP or living long-term in an institution any time during measurement year. Exclude these patients from the eligible population.

Step 3. Determine numerator: the number of patients who received one or more mammograms any time on or between October 1 two years prior to the measurement year and December 31 of the measurement year.

Step 4. Calculate the rate. Step 1. Determine the eligible population: identify women 52-74 years of age by the end of the measurement year.

Step 2. Search for an exclusion: history of bilateral mastectomy; or use of hospice services during the measurement year; or patients 65 and older who are enrolled in an institutional SNP or living long-term in an institution any time during measurement year. Exclude these patients from the eligible population.

Step 3. Determine numerator: the number of patients who received one or more mammograms any time on or between October 1 two years prior to the measurement year and December 31 of the measurement year.

Step 4. Calculate the rate.

Submission items

0508: Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

5.1 Identified measures:

5a.1 Are specs completely harmonized? No

5a.2 If not completely harmonized, identify difference, rationale, impact: The OP-9 measure is calculated using administrative claims data. The period of data collection for OP-9 is only 45 days, and most code 3 recall is 90 or 180 days.

5b.1 If competing, why superior or rationale for additive value: There are no competing measures (conceptually both the same measure focus and same target population)

2372: Breast Cancer Screening

5.1 Identified measures: 0508 : Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms

0509 : Diagnostic Imaging: Reminder System for Screening Mammograms

5a.1 Are specs completely harmonized? Yes

5a.2 If not completely harmonized, identify difference, rationale, impact: Both related measures have a different focus than our health plan screening measure. NQF #0509 Reminder System for Mammograms is intended to encourage implementation of reminder systems for future mammograms. NQF #0508 Inappropriate Use of "Probably Benign" Assessment Category focuses on accurate documentation of mammogram results. Both measures are also specified at the clinician level rather than the health plan level.

5b.1 If competing, why superior or rationale for additive value: N/A

Appendix F: Pre-Evaluation Comments

Pre-meeting commenting closed on June 19, 2020. As of that date, no comments were submitted.

National Quality Forum
1099 14th Street NW, Suite 500
Washington, DC 20005
http://www.qualityforum.org