



Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

Brief Measure Information

NQF #: 2372

Corresponding Measures:

Measure Title: Breast Cancer Screening

Measure Steward: National Committee for Quality Assurance

sp.02. Brief Description of Measure: Percentage of women 50-74 years of age who had a mammogram to screen for breast cancer

1b.01. Developer Rationale: This measure assesses screening for breast cancer using mammography, which can prevent or detect early breast cancer, as well as reduce deaths from breast cancer. Early detection of breast cancer by mammography may also allow for a greater range of treatment options, including less-aggressive surgery and less-invasive therapy.

sp.12. Numerator Statement: Women who received a mammogram to screen for breast cancer.

sp.14. Denominator Statement: Women 50-74 years of age.

sp.16. Denominator Exclusions: 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.

Measure Type: Process

sp.28. Data Source:

Electronic Health Data

Claims

sp.07. Level of Analysis:

Integrated Delivery System

Health Plan

IF Endorsement Maintenance – Original Endorsement Date: 2014-09-18 01:15 PM

Most Recent Endorsement Date: 10/25/2018 2:36:09 PM

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:

1. Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria

1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.

[Response Begins]

Yes

[Response Ends]

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01. Provide a logic model.

Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

[Response Begins]

[Response Ends]

1a.02. Select the type of source for the systematic review of the body of evidence that supports the performance measure.

A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data.

[Response Begins]

[Response Ends]

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, add additional tables by clicking "Add" after the final question in the group.

Evidence - Systematic Reviews Table (Repeatable)

Group 1 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]

[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]

[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]

[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]

[Response Ends]

1a.13. If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.

[Response Begins]

[Response Ends]

1a.14. Briefly synthesize the evidence that supports the measure.

[Response Begins]

[Response Ends]

1a.15. Detail the process used to identify the evidence.

[Response Begins]

[Response Ends]

1a.16. Provide the citation(s) for the evidence.

[Response Begins]

[Response Ends]

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

This measure assesses screening for breast cancer using mammography, which can prevent or detect early breast cancer, as well as reduce deaths from breast cancer. Early detection of breast cancer by mammography may also allow for a greater range of treatment options, including less-aggressive surgery and less-invasive therapy.

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

The following health-plan level data are collected through HEDIS and reflect the most recent years of measurement for this measure. Performance data are summarized at the health plan level and summarized by mean, standard deviation, minimum health plan performance, maximum health plan performance and

performance at 10th, 25th, 50th, 75th, and 90th percentile. Data are stratified by year and product line (i.e. commercial, Medicaid, Medicare).

Commercial HMO/PPO Rate

YEAR	N	MEAN	STDEV	MIN	10TH	25TH	50TH	75TH	90TH	MAX
2015	410	71.9%	5.9%	52.0%	64.8%	68.1%	71.1%	75.8%	80.5%	89.6%
2016	426	71.4%	6.5%	37.2%	64.5%	67.7%	71.0%	75.4%	80.2%	88.2%
2017	420	71.4%	7.0%	13.2%	64.0%	68.0%	71.2%	75.7%	79.8%	88.0%

Medicaid HMO Rate

YEAR	N	MEAN	STDEV	MIN	10TH	25TH	50TH	75TH	90TH	MAX
2015	170	58.77%	09.75%	36.70%	45.99%	51.59%	58.37%	66.02%	71.32%	87.88%
2016	201	58.51%	10.09%	17.78%	47.38%	52.28%	58.10%	65.06%	71.44%	88.51%
2017	257	58.87%	09.24%	30.56%	48.00%	52.71%	59.02%	65.51%	70.29%	87.92%

Medicare HMO/PPO Rate

YEAR	N	MEAN	STDEV	MIN	10TH	25TH	50TH	75TH	90TH	MAX
2015	425	71.0%	11.3%	09.8%	60.5%	66.8%	71.7%	78.4%	82.9%	92.1%
2016	405	72.4%	09.8%	14.3%	62.0%	67.4%	72.7%	79.2%	83.3%	91.9%
2017	431	72.2%	09.6%	18.4%	61.4%	67.0%	73.0%	78.8%	82.9%	91.1%

In 2016, HEDIS measures covered 114.2 million commercial health plan beneficiaries, 47 million Medicaid beneficiaries and 17.6 million Medicare beneficiaries. Below is a description of the denominator for this measure. It includes the number of health plans included in HEDIS data collection and the mean eligible population for the measure across health plans.

Breast Cancer Screening – commercial

YEAR	N Plans	Median Denominator Size per plan
2015	410	7,740
2016	426	7,510
2017	420	7,552

Breast Cancer Screening – Medicaid

YEAR	N Plans	Median Denominator Size per plan
2015	172	1,642
2016	202	1,419
2017	258	2,439

Breast Cancer Screening – Medicare

YEAR	N Plans	Median Denominator Size per plan
2015	425	2,173
2016	405	2,297
2017	431	1,890

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.

[Response Begins]

N/A

[Response Ends]

1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., “topped out”, disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

HEDIS data are stratified by type of insurance (e.g. Commercial, Medicaid, Medicare). While not specified in the measure, health plans can stratify by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities. The Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, the NCQA Multicultural Health Care Distinction Program outlines standards for collecting, storing and using race/ethnicity and language data to assess health care disparities. Starting in 2019, Medicare Advantage plans will report this measure stratified by low-income subsidy/dual eligibility and disability status, which are proxies for low socioeconomic status.

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.

[Response Begins]

Studies have identified disparities in breast cancer screening based on race, ethnicity, education and income. One study found that mammography use in 2006 was 65 percent among white women and 59 percent among black women (Njai et al 2011). Additionally, African American women are more likely than white women to have longer intervals between screening mammograms, which may lead to an increase in later-stage cancer diagnoses (CDC 2012). Between 2010 and 2014, breast cancer mortality for African American women was 41 percent higher than white women (Richardson et al 2016); one potential contributing factor to this health disparity is access to mammography screening services (Rust et al 2015). National survey data also show that women who have attained lower degrees of education, lack health insurance coverage or have lower socioeconomic status are less likely to have recently had a mammogram (National Center for Health Statistics 2015).

National Center for Health Statistics. 2015. “Health, United States, 2015: With Special Feature on Racial and Ethnic Health Disparities”. <http://www.cdc.gov/nchs/data/abus/abus15.pdf#070> (Accessed Mar 13, 2018).

Njai, R., Siegel, P. Z., Miller, J. W., & Liao, Y. 2011. “Misclassification of Survey Responses and Black-White Disparity in Mammography Use, Behavioral Risk Factor Surveillance System, 1995-2006.” Preventing Chronic Disease 8(3), A59.

Richardson, L.C., Henley, S.J., Miller, J.W., Massetti, G., and Thomas, C.C. 2016. “Patterns and Trends in Age-Specific Black-White Differences in Breast Cancer Incidence and Mortality – United States, 1999–2014.” Morbidity and Mortality Weekly Report (MMWR) 65(40); 1093-1098. (November 30, 2016) <http://www.cdc.gov/mmwr/volumes/65/wr/mm6540a1.htm>.

Rust, G., Zhang, S., Malhotra, K., Reese, L., McRoy, L., Baltrus, P., Caplan, L. and Levine, R. 2015. “Paths to Health

Equity – Local Area Variation in Progress Toward Eliminating Breast Cancer Mortality Disparities, 1990–2009.”
Cancer 121(16): 2765-2774. (December 1, 2016) doi: 10.1002/cncr.29405.

[Response Ends]

2. Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.

spma.01. Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.

[Response Begins]

Yes

[Response Ends]

spma.02. Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

For example, specifications may have been updated based on suggestions from a previous NQF CDP review.

[Response Begins]

Since the last NQF review, digital breast tomosynthesis was added as an acceptable breast cancer screening method in order to account for the use of this method by women with clinical indications. This change was reviewed by stakeholder groups, vetted through a public comment period, and approved by our committees.

[Response Ends]

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see [What Good Looks Like](#)).

[Response Begins]

Breast Cancer Screening

[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]

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

[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]

Cancer

Cancer: Breast

[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]

Primary Prevention

Screening

[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Health Plan

Integrated Delivery System

[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.

[Response Begins]

Outpatient Services

[Response Ends]

sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.

Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".

[Response Begins]

N/A

[Response Ends]

sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, [contact staff](#). Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

[Response Begins]

No data dictionary/code table – all information provided in the submission form

[Response Ends]

sp.13. State the numerator.

Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).

DO NOT include the rationale for the measure.

[Response Begins]

Women who received a mammogram to screen for breast cancer.

[Response Ends]

sp.14. Provide details needed to calculate the numerator.

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

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.

[Response Ends]

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

Women 50-74 years of age.

[Response Ends]

sp.16. Provide details needed to calculate the denominator.

All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

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.

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

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.

[Response Ends]

sp.18. Provide details needed to calculate the denominator exclusions.

All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

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.

[Response Ends]

sp.19. Provide all information required to stratify the measure results, if necessary.

Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.

[Response Begins]

N/A

[Response Ends]

sp.20. Is this measure adjusted for socioeconomic status (SES)?

[Response Begins]

[Response Ends]

sp.21. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

[Response Begins]

No risk adjustment or risk stratification

[Response Ends]

sp.22. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

sp.23. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score

[Response Begins]

Better quality = Higher score

[Response Ends]

sp.24. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]

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.

[Response Ends]

sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

- Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.
- The sample should represent the variety of entities whose performance will be measured. The [2010 Measure Testing Task Force](#) recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.

- *The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.*
- *When possible, units of measurement and patients within units should be randomly selected.*

[Response Begins]

N/A

[Response Ends]

sp.30. Select only the data sources for which the measure is specified.

[Response Begins]

Claims

Electronic Health Data

[Response Ends]

sp.31. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]

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.

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

2ma.02. Indicate whether additional empirical validity testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?

[Response Begins]

[Response Ends]

2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

[Response Begins]

[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.

- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the [Submitting Standards webpage](#).
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the [2021 Measure Evaluation Criteria and Guidance](#).

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measure scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Risk factors that influence outcomes should not be specified as exclusions.

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of \$25 in cost for an episode of care (e.g., \$5,000 v. \$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]

[Response Ends]

2a.02. If an existing dataset was used, identify the specific dataset.

The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

[Response Begins]

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

[Response Ends]

2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).

Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.

[Response Begins]

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

[Response Ends]

2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.

[Response Begins]

[Response Ends]

2a.08. List the social risk factors that were available and analyzed.

For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

[Response Begins]

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter “see validity testing section of data elements”; and enter “N/A” for 2a.11 and 2a.12.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.

[Response Begins]

[Response Ends]

2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.

Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.

[Response Begins]

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, [NQF Measure Evaluation Criteria](#)).

[Response Begins]

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]

[Response Ends]

2b.01. Select the level of validity testing that was conducted.

[Response Begins]

[Response Ends]

2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.

Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.

[Response Begins]

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]

[Response Ends]

2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)

[Response Begins]

[Response Ends]

2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.

Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.

[Response Begins]

[Response Ends]

2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]

[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.

[Response Begins]

[Response Ends]

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]

[Response Ends]

2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.

Describe the steps—do not just name a method. Indicate what statistical analysis was used.

[Response Begins]

[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]

[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]

[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?

[Response Begins]

[Response Ends]

2b.17. Provide the statistical results from testing exclusions.

Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.

[Response Begins]

[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.

[Response Begins]

[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]

[Response Ends]

2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]

[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]

[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of $p < 0.10$ or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]

[Response Ends]

2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter “N/A” for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]

[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]

[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]

[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]

[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]

[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]

[Response Ends]

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

ALL data elements are in defined fields in electronic claims

[Response Ends]

3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.

[Response Begins]

N/A

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

[Response Ends]

3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

[Response Begins]

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, algorithm),

Attach the fee schedule here, if applicable.

[Response Begins]

Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

[Response Ends]

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

4a.01. Check all current uses. For each current use checked, please provide:

Name of program and sponsor

URL

Purpose

Geographic area and number and percentage of accountable entities and patients included

Level of measurement and setting

[Response Begins]

Public Reporting

Payment Program

Regulatory and Accreditation Programs

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]

N/A

[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

[Response Begins]

N/A

[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.

[Response Begins]

During development, NCQA receives input from those reporting and using measures through several multistakeholder advisory panels and a broad public comment posting. For this particular measure, the clinical advisory panel included several representatives from health plans and users such as federal policymakers and consumers. We also sought input from our standing Technical Measurement Advisory Panel, which includes representatives from health plans and other users and advises NCQA on feasibility and other potential implementation issues. During implementation, health plans that report HEDIS calculate their rates and know their performance when submitting to NCQA. NCQA publicly reports rates across all plans and also creates benchmarks in order to help plans understand how they perform relative to other plans. Public reporting and benchmarking are effective quality improvement methods.

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]

NCQA publishes HEDIS results annually in our Quality Compass tool. NCQA also presents data at various conferences and webinars. For example, at the annual HEDIS Update and Best Practices Conference, NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly. NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System, as described in Section 3c.1.

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]

NCQA measures are evaluated regularly using a consensus-based process to consider input from multiple stakeholders, including but not limited to entities being measured. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System. This information enables NCQA to comprehensively assess a measure's adherence to the HEDIS Desirable Attributes of Relevance, Scientific Soundness and Feasibility.

[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]

During development and reevaluation, those reporting and using measures reported the measure continues to be relevant and important for quality improvement and accountability. Questions received through the Policy Clarification Support system have generally sought clarification about the mammography screening methods that satisfy the measure numerator. During a recent public comment session, a majority of comments from measured entities supported updates to the measure to align with the latest clinical recommendations.

[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]

This measure has been deemed a priority measure by NCQA and other entities, as illustrated by its use in programs such as the Medicare Advantage Star Rating program and the Medicaid Adult Core Set.

[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]

During the measure's last major update, feedback obtained through the mechanisms described in 4a2.2.1 informed how we revised the measure to include digital breast tomosynthesis as a new screening method.

[Response Ends]

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

[Response Begins]

From 2011 to 2017, average performance rates increased by about six percentage points for Medicare and Medicaid plans and were steady for commercial plans. Over the past three years, average performance rates slightly increased for Medicare plans and were stable for Commercial and Medicaid plans. In 2017, average performance was about 71% for Commercial and Medicare plans, and 59% for Medicaid plans.

There continues to be variation between the 10th and 90th percentile, suggesting room for improvement. In 2017, commercial plans in the 10th percentile had a rate of 64% compared to 80% for plans in the 90th percentile; Medicare plans in the 10th percentile had a rate of 61% compared to 83% for plans in the 90th percentile; and Medicaid plans in the 10th percentile had a rate of 48% compared to 70% for plans in the 90th percentile.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]

One potential unintended consequence of the Breast Cancer Screening measure is too-frequent screening. The U.S. Preventive Services Task Force recommends biennial screening in women age 50-74. Feedback from our advisory panel indicated that, in an effort to meet the two-year requirement, women often are encouraged to seek screening earlier than the two-year mark. In order to address potential over-screening, NCQA adjusted the numerator time frame to 27 months, providing a three-month leeway to account for the logistics of scheduling and receiving a mammogram.

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]

Benefits from implementation of the measure were as expected.

[Response Ends]

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

N/A

[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

Yes

[Response Ends]

5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

[Response Begins]

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.

[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]

N/A

[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.:

No appendix

Contact Information

Measure Steward (Intellectual Property Owner): National Committee for Quality Assurance

Measure Steward Point of Contact: Rehm, Bob, nqf@ncqa.org

Rehm, Bob, rehm@ncqa.org

Wade, Brittany, wade@ncqa.org

Measure Developer if different from Measure Steward: National Committee for Quality Assurance

Measure Developer Point(s) of Contact: Rehm, Bob, nqf@ncqa.org

Wade, Brittany, wade@ncqa.org

Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]

No appendix

[Response Ends]

2. List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

[Response Begins]

The NCQA Breast Cancer Screening Measurement Advisory Panels advised NCQA during measure development. They evaluated the way staff specified the measure, reviewed field test results, and assessed NCQA's overall desirable attributes of Relevance, Scientific Soundness, and Feasibility. The advisory panel consisted of a balanced group of experts. In addition to this advisory panel, we vetted the measure with a host of other stakeholders, as is our process. Thus, our measures are the result of consensus from a broad and diverse group of stakeholders.

2014 BREAST CANCER SCREENING MEASUREMENT ADVISORY PANEL MEMBERS

Kathy Coltin, Harvard Pilgrim Health Care
Laura Esserman, University of California, San Francisco
Lisa Latts, formally at WellPoint, Inc.
Nancy Lee, U.S. Department of Health & Human Services
Dorothy Mann, UW School of Medicine and School of Public Health and Community Medicine
Melissa McNeil, University of Pittsburgh
Ellen Stovall, National Coalition for Cancer Survivorship
Richard Wender, Thomas Jefferson University

2017 BREAST CANCER SCREENING MEASUREMENT ADVISORY PANEL MEMBERS

Joanne Armstrong, Aetna
Laura Esserman, University of California, San Francisco
Sandra Finestone, Association of Cancer Patient Educators
David Larsen, Intermountain Healthcare
Melissa McNeil, University of Pittsburgh, UPMC
Robert Smith, American Cancer Society

2017 GERIATRIC MEASUREMENT ADVISORY PANEL MEMBERS

Arlene Bierman, AHRQ
Patricia Bomba, Excellus BlueCross BlueShield
Jennie Chin Hansen, American Geriatrics Society (Retired)
Joyce Dubow, Public Member
Peter Hollmann, Brown University
Steven Phillips, Geriatric Specialty Care
Wade Aubry, UCSF Institute for Health Policy
Jane Sung, AARP
Eric Tangalos, Mayo Clinic
Dirk Wales, Cigna HealthSpring
Neil Wenger, UCLA

Nicole Brandt, UMD Pharmacy
Karen Nichols, Amerihealth Caritas
Gustavo Ferrer, Aventura Hospital
Jeff Kelman, CMS
Joan Weiss, HHS

2017 COMMITTEE ON PERFORMANCE MEASUREMENT

Bruce Bagley, Independent Consultant
Andrew Baskin, Aetna
Jonathan D. Darer, Medicalis
Helen Darling, Strategic Advisor on Health Benefits & Health Care
Kate Goodrich, Centers for Medicare and Medicaid Services
David Grossman, Kaiser Permanente
Christine S. Hunter, US Office of Personnel Management
Jeffrey Kelman, Centers for Medicare & Medicaid Services
Nancy Lane, Vanderbilt University Medical Center
Bernadette Loftus, The Permanente Medical Group
Adrienne Mims, Alliant Quality
Amanda Parsons, Montefiore Health System
Eric C. Schneider, The Commonwealth Fund
Marcus Thygeson, Blue Shield of California
JoAnn Volk, Georgetown University Center on Health Insurance Reforms

2017 TECHNICAL MEASUREMENT ADVISORY PANEL MEMBERS

Andy Amster, Kaiser Permanente
Jennifer Brudnicki, Geisinger Health Plan
Lindsay Cogan, New York State Department of Health
Kathy Coltin, Independent Consultant
Mike Farina, MVP Healthcare
Marissa Finn, CIGNA HealthCare
Scott Fox, Independence Blue Cross
Carlos Hernandez, CenCal Health
Harmon Jordan, Westat
Virginia Raney, Center for Medicaid and CHIP Services
Lynne Rothney-Kozlak, Rothney-Kozlak Consulting, LLC
Laurie Spoll, Aetna

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]

[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]

[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]

Approximately every 3 years, sooner if the clinical guidelines have changed significantly.

[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]

[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate “N/A”.

[Response Begins]

The performance measures and specifications were developed by and are owned by the National Committee for Quality Assurance (“NCQA”). The performance measures and specifications are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties, or endorsement about the quality of any organization or physician that uses or reports performance measures and NCQA has no liability to anyone who relies on such measures or specifications. NCQA holds a copyright in these materials and can rescind or alter these materials at any time. These materials may not be modified by anyone other than NCQA. Anyone desiring to use or reproduce the materials without modification for an internal, quality improvement non-commercial purpose may do so without obtaining any approval from NCQA. All other uses, including a commercial use and/or external reproduction, distribution and publication must be approved by NCQA and are subject to a license at the discretion of NCQA.

©2018 NCQA, all rights reserved.

Limited proprietary coding is contained in the measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. NCQA disclaims all liability for use or accuracy of any coding contained in the specifications.

Content reproduced with permission from HEDIS, Volume 2: Technical Specifications for Health Plans. To purchase copies of this publication, including the full measures and specifications, contact NCQA Customer Support at 888-275-7585 or visit www.ncqa.org/publications.

[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate “N/A”.

[Response Begins]

These performance Measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.

THE MEASURES AND SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate “N/A”.

[Response Begins]

NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties, or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance, or otherwise modify the measures, and shall not disassemble, recompile, or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA. © 2018 by the National Committee for Quality Assurance

[Response Ends]