NATIONAL QUALITY FORUM

Measure Submission and Evaluation Worksheet 5.0

This form contains the information submitted by measure developers/stewards, organized according to NQF's measure evaluation criteria and process. The evaluation criteria, evaluation guidance documents, and a blank online submission form are available on the submitting standards web page.

NQF #: 0031   **NQF Project:** Cancer Project
(for Endorsement Maintenance Review) **Original Endorsement Date:** Aug 10, 2009 **Most Recent Endorsement Date:** Aug 10, 2009 **Last Updated Date:** Oct 01, 2014

BRIEF MEASURE INFORMATION
De.1 Measure Title: Breast Cancer Screening
Co.1.1 Measure Steward: National Committee for Quality Assurance
De.2 Brief Description of Measure: Percentage of women 40-69 years of age who had a mammogram to screen for breast cancer
2a1.1 Numerator Statement: One or more mammograms during the measurement year or the year prior to the measurement year
2a1.4 Denominator Statement: Women 42–69 years of age as of Dec 31 of the measurement year (note: this denominator statement captures women age 40-69 years)
2a1.8 Denominator Exclusions: Exclusion: Women who had a bilateral mastectomy or for whom there is evidence of two unilateral mastectomies. Look for evidence of a bilateral mastectomy as far back as possible int he member´s history thorugh Dec 31 of the measurement year.
1.1 Measure Type: Process **2a1. 25-26 Data Source:** Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record **2a1.33 Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Health Plan, Population : State **1.2-1.4 Is this measure paired with another measure?** No **De.3 If included in a composite, please identify the composite measure** (*title and NQF number if endorsed*)**:** NA

STAFF NOTES (*issues or questions regarding any criteria*)
Comments on Conditions for Consideration:
Is the measure untested? Yes⊙ No⊙   If untested, explain how it meets criteria for consideration for time-limited endorsement:
1a. Specific national health goal/priority identified by DHHS or NPP addressed by the measure (*check De.5*)**:** **5. Similar/related endorsed or submitted measures** (*check 5.1*)**:** **Other Criteria:**
Staff Reviewer Name(s):

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.

Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1a. High Impact: H⬤ M⬤ L⬤ I ⬤
(*The measure directly addresses a specific national health goal/priority identified by DHHS or NPP, or some other high impact aspect of healthcare.*)

De.4 Subject/Topic Areas (*Check all the areas that apply*): Cancer, Cancer : Breast, Cancer : Screening, Prevention
De.5 Cross Cutting Areas (*Check all the areas that apply*): Prevention

1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, A leading cause of morbidity/mortality, Severity of illness

1a.2 If "Other," please describe:

1a.3 Summary of Evidence of High Impact (*Provide epidemiologic or resource use data*)**:**
Breast cancer is one of the most common types of cancers, accounting for a quarter of all the new cancer diagnoses in American women.1 Breast cancer is the second top cause of cancer deaths in women (after lung cancer) with nearly 40,000 estimated deaths in 2010.2

1a.4 Citations for Evidence of High Impact cited in 1a.3: 1. BreastCancer.org. 2011. U.S Breast Cancer Statistics. http://www.breastcancer.org/symptoms/understand_bc/statistics.jsp (June 10, 2011).

2. American Cancer Society. 2011. Cancer Facts & Figures 2011. http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-026238.pdf (May 29, 2011).

3. American Cancer Society. 2010. Breast cancer facts and figures 2010-2011. http://www.cancer.org/acs/groups/content/@nho/documents/document/f861009final90809pdf.pdf (May 29, 2011).

4. National Cancer Institute. 2010. Breast Cancer Screening. http://www.cancer.gov/cancertopics/pdq/screening/breast/healthprofessional (May 29, 2011).

5. BreastCancer.org. 2011. Facts and Figures about Breast Cancer http://www.breastcancer.org/about_us/press_room/facts_figures.jsp (June 10, 2011)

6. National Business Group on Health. 2011. Pathways to Managing Cancer in the Workplace. http://www.businessgrouphealth.org/pdfs/FINAL_Pathways_Managing_Cancer_2011.pdf (June 10, 2011).

7. Screening for Breast Cancer: An Update for the U.S. Preventive Services Task Force. 2009. Annals of Internal Medicine.151:738-47. http://www.annals.org/content/151/10/727.full.pdf+html (May 29, 2011).

8. Nelson, H.D., Tyne, K., Naik, A., Bougatsos, C., Chan, B.K., Humphrey, L. Screening for Breast Cancer: An Update for the U.S. Preventive Services Task Force. Ann Intern Med. 2009;151:727-737.

1b. Opportunity for Improvement: H⬤ M⬤ L⬤ I ⬤
(*There is a demonstrated performance gap - variability or overall less than optimal performance*)

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:
The intent of the measure is to improve secondary prevention of breast cancer in order to catch disease when it is early and more amenable to treatment. The five-year survival rate for women who are diagnosed early is 98 percent compared to a late-diagnosed breast cancer survival rate of only 23 percent. Studies suggest that, on average, mammography will detect about 80-90 percent of breast cancers in women without symptoms. In addition, early detection of breast cancer by mammography may allow for a greater range of treatment options, including less-aggressive surgery and less-invasive therapy.

1b.2 Summary of Data Demonstrating Performance Gap *(Variation or overall less than optimal performance across providers)*: **[For _Maintenance_ –** *Descriptive statistics for performance results for this measure - distribution of scores for measured entities by quartile/decile, mean, median, SD, min, max, etc.*]
HEDIS Health Plan measure rates suggest there is room for improvement in breast cancer screening, with average rates of 71% for commercial plans; 69% for Medicare plans; and 52% for Medicaid plans in 2009.

Commercial Health Plan Rates

Year	2009	2008	2007
N	244	253	258
MEAN	71.3	70.2	69.1
STDEV	6.08	6.18	5.94
STDERR	0.39	0.39	0.37
MIN	55.1	47.5	52.2
MAX	90	85.1	84.4
P10	64.2	62.7	61.9
P25	67.0	66.2	64.9
P50	70.7	70	68.5
P75	75.3	74.2	73.5
P90	80.1	78.7	78.2

Medicare Health Plans

Year	2009	2008	2007
N	290	255	229
MEAN	69.3	68	67.3
STDEV	10.0	11.2	11.7
STDERR	0.59	0.7	0.77
MIN	37.8	13.9	19
MAX	94.7	95.2	90.4
P10	55.7	53.2	52.3
P25	62.5	60.5	60.7
P50	69.8	67.8	68
P75	75.6	75.2	75.1
P90	82.7	82.9	83.5

Medicaid Health Plans

Year	2009	2008	2007
N	144	137	138
MEAN	52.4	50.8	49.8
STDEV	10.2	10.3	9.36
STDERR	0.85	0.88	0.8
MIN	25.2	18.5	9.38
MAX	78.4	81.7	77.1
P10	39.8	38.6	38.8
P25	46.2	45	44.3

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

P50		52	50.5	50
P75		59.6	57.4	56.1
P90		63.8	63	61.2

1b.3 Citations for Data on Performance Gap: [*For Maintenance – Description of the data or sample for measure results reported in 1b.2 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included***]**
Data shown are performance rates from health plans reporting the HEDIS health plan measure. There were 2108 plan submissions for this measure.

1b.4 Summary of Data on Disparities by Population Group: [*For Maintenance –Descriptive statistics for performance results for this measure by population group***]**
The latest reports show that African American and white women now have the same rate of mammography use. In 2008, 68% of African American women and 68% of white women had a mammogram within the past two years. Hispanic/Latina, Asian American and Native American have lower rates of breast cancer screening compared to African American and white women (62% of Hispanic/Latina women, 65%of Asian American women and 55% of Native American women had a mammogram in the past two years).

The reasons behind these differences in mammography rates are unclear. Women have reported that costs, lack of insurance, and poor access to care or a usual health care provider are barriers to mammography.

1b.5 Citations for Data on Disparities Cited in 1b.4: [*For Maintenance – Description of the data or sample for measure results reported in 1b.4 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included***]**
Centers for Disease Control and Prevention. 2010. Vital Signs. http://www.cdc.gov/VitalSigns/pdf/2010-07-vitalsigns.pdf (July 8).

1c. Evidence (*Measure focus is a health outcome OR meets the criteria for quantity, quality, consistency of the body of evidence.*)
Is the measure focus a health outcome?   Yes⊘ No⦿   If not a health outcome, rate the body of evidence.

Quantity: H⦿ M⦿ L⦿ I ⦿   Quality: H⦿ M⦿ L⦿ I ⦿   Consistency: H⦿ M⦿ L⦿ I ⦿

Quantity	Quality	Consistency	Does the measure pass subcriterion1c?
M-H	**M-H**	**M-H**	**Yes⦿**
L	**M-H**	**M**	**Yes⦿** IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise **No⦿**
M-H	**L**	**M-H**	**Yes⦿** IF potential benefits to patients clearly outweigh potential harms: otherwise **No⦿**
L-M-H	**L-M-H**	**L**	**No ⦿**

Health outcome – rationale supports relationship to at least one healthcare structure, process, intervention, or service	**Does the measure pass subcriterion1c?** **Yes⦿** IF rationale supports relationship

1c.1 Structure-Process-Outcome Relationship (*Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome*)**:**
This measure focuses on a process of care (screening). Research suggests screening leads to earlier detection of breast cancer, which leads to more effective treatment and lower mortality rates.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
This measure does not differ in focus from the body of evidence.

1c.5 Quantity of Studies in the <u>Body of Evidence</u> *(Total number of studies, not articles)*: See USPSTF guideline report

1c.6 Quality of <u>Body of Evidence</u> (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events)*: High

1c.7 Consistency of Results <u>across Studies</u> *(Summarize the consistency of the magnitude and direction of the effect)*: Consistent within guidelines

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
The USPSTF determined there was a positive net benefit for breast cancer screening.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: USPSTF

1c.11 System Used for Grading the Body of Evidence: USPSTF

1c.12 If other, identify and describe the grading scale with definitions:

1c.13 Grade Assigned to the Body of Evidence: The USPSTF recommends biennial screening mammography for women aged 50 to 74 years. Grade B.

1c.14 Summary of Controversy/Contradictory Evidence: Major guidelines differ regarding the age at which women should begin screening. The USPSTF's 2009 guideline recommends raising the age at which women begin screening from 40 to 50 years (a change from their 2002 guideline). Other guidelines, such as the American Cancer Society and the American Congress of Obstetricians and Gyncologists, recommend screening begin earlier.

Issues related to the controversy over the age at which to begin screening involve the trade-offs between harms and benefits. The USPSTF concluded that the overall harms associated with mammography are moderate for every age group considered, although the main components of the harms shift over time. Although false-positive test results, overdiagnosis, and unnecessary earlier treatment are problems for all age groups, false-positive results are more common for women aged 40 to 49 years, whereas overdiagnosis is a greater concern for women in the older age groups.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

Nelson, H.D., Tyne, K., Naik, A., Bougatsos, C., Chan, B.K., Humphrey, L. Screening for Breast Cancer: An Update for the U.S. Preventive Services Task Force. Ann Intern Med. 2009;151:727-737.

Kolata, G. Behind Cancer Guidelines, Quest for Data. New York Times, November 23 2009.

1c.16 **Quote verbatim, <u>the specific guideline recommendation</u>** *(Including guideline # and/or page #)*:
U.S. Preventive Services Task Force (2009)
Grade: B recommendation. The USPSTF recommends biennial screening mammography for women aged 50 to 74 years.
Grade: C recommendation. The decision to start regular, biennial screening mammography before the age of 50 years should be an individual one and take patient context into account, including the patient´s values regarding specific benefits and harms.
Grade: I Statement. The USPSTF concludes that the current evidence is insufficient to assess the additional benefits and harms of screening mammography in women 75 years or older.
Grade: D recommendation. The USPSTF recommends against teaching breast self-examination (BSE).
Grade: I Statement. The USPSTF concludes that the current evidence is insufficient to assess the additional benefits and harms of clinical breast examination (CBE) beyond screening mammography in women 40 years or older.
Grade: I Statement. The USPSTF concludes that the current evidence is insufficient to assess the additional benefits and harms of either digital mammography or magnetic resonance imaging (MRI) instead of film mammography as screening modalities for breast cancer.

American Academy of Family Physicians
Endorses the USPSTF recommendation

American Cancer Society (2009)
Recommends annual screening using mammography and clinical breast examination for all women beginning at age 40.

American College of Radiology (2009)
Recommends annual screening using mammography and clinical breast examination for all women beginning at age 40.

American Congress of Obstetricians and Gynecologists (2009)
Recommends:
• Screening mammography every 1-2 years for women aged 40-49 years
• Screening mammography every year for women age 50 or older
• BSE; BSE has the potential to detect palpable breast cancer and can be recommended.

1c.17 **Clinical Practice Guideline Citation:** American Cancer Society (ACS). (1) ACS guidelines for breast cancer screening: update 2003. (2) American Cancer Society Guideline for breast screening with MRI as an adjunct to mammography (2007). CA Cancer J Clin 2007 Mar-Apr;57(2):75-89.

American Academy of Family Physicians
http://www.aafp.org/online/en/home/clinical/exam/a-e.html

American College of Obstetricians and Gynecologists (ACOG). Breast cancer screening. Washington (DC): American College of Obstetricians and Gynecologists (ACOG); 2003 Apr. 12 p. (ACOG practice bulletin; no.42).

American College of Physicians (ACP). Screening mammography for women 40 to 49 years of age: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2007 Apr 3;146(7):511-5.

U.S. Preventive Services Task Force (USPSTF). 1) Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. 2) December 2009 addendum. Ann Intern Med 2009 Nov 17;151(10):716-726.

1c.18 National Guideline Clearinghouse or other URL:
http://www.guideline.gov/content.aspx?id=3990#Section427

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: USPSTF

1c.21 System Used for Grading the Strength of Guideline Recommendation: USPSTF

1c.22 If other, identify and describe the grading scale with definitions:

1c.23 Grade Assigned to the Recommendation: The USPSTF recommends biennial screening mammography for women aged 50 to 74 years. Grade B.

1c.24 Rationale for Using this Guideline Over Others: NCQA bases its measures on clinical guidelines whenever possible. When NCQA developed this measure, we based the measure primarily on the USPSTF 2002 recommendation for screening women starting at age 40. Since then, the USPSTF has updated its guideline, noting that screening women under age 50 should be an individual decision taking patient context and values into account. Many other entities still recommend screening begin earlier. Due to the conflict in guidelines, NCQA kept the measure at age 40 and is planning to begin a full re-evaluation in spring of 2012.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: High   **1c.26** Quality: High **1c.27** Consistency: Moderate
1c.28 Attach evidence submission form:
1c.29 Attach appendix for supplemental materials:

Was the threshold criterion, *Importance to Measure and Report*, met?
(*1a & 1b must be rated moderate or high and 1c yes*)   Yes◉ No◉
Provide rationale based on specific subcriteria:

For a new measure if the Committee votes NO, then STOP.
For a measure undergoing endorsement maintenance, if the Committee votes NO because of 1b. (no opportunity for improvement), it may be considered for continued endorsement and all criteria need to be evaluated.

2. RELIABILITY & VALIDITY - 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. **(evaluation criteria)**
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 field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

S.1 Measure Web Page (*In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained*). Do you have a web page where current detailed specifications for this measure can be obtained? No

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

S.2 **If yes, provide web page URL:** TBD

2a. RELIABILITY. Precise Specifications and Reliability Testing: H⊙ M⊙ L⊙ I ⊙

2a1. Precise Measure Specifications. (*The measure specifications precise and unambiguous.*)

2a1.1 Numerator Statement (*Brief, narrative description of the measure focus or what is being measured about the target population, e.g., cases from the target population with the target process, condition, event, or outcome*)**:**
One or more mammograms during the measurement year or the year prior to the measurement year

2a1.2 Numerator Time Window (*The time period in which the target process, condition, event, or outcome is eligible for inclusion*)**:**
2 years

2a1.3 Numerator Details (*All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, codes with descriptors, and/or specific data collection items/responses*)**:**
A woman had a mammogram if a submitted claim/encounter contains any of the following codes.
Codes to Identify Breast Cancer Screening:
CPT: 77055-77057
HCPCS: G0202, G0204, G0206
ICD-9-CM Procedure: 87.36, 87.37
UB Revenue: 0401, 0403

Note: The purpose of this measure is to evaluate primary screening. Do not count biopsies, breast ultrasounds or MRIs because they are not appropriate methods for primary breast cancer screening.

2a1.4 Denominator Statement (*Brief, narrative description of the target population being measured*)**:**
Women 42–69 years of age as of Dec 31 of the measurement year (note: this denominator statement captures women age 40-69 years)

2a1.5 Target Population Category (*Check all the populations for which the measure is specified and tested if any*)**:** Senior Care

2a1.6 Denominator Time Window (*The time period in which cases are eligible for inclusion*)**:**
1 Year

2a1.7 Denominator Details (*All information required to identify and calculate the target population/denominator such as definitions, codes with descriptors, and/or specific data collection items/responses*)**:**
Product lines: Commercial, Medicaid, Medicare
Ages: Women 42-69 years as of December 31 of the measurement year
Continuous Enrollment: The measurement year and the year prior to the measurement year
Allowable gap: No more than one gap of enrollment of up to 45 days during each year of continuous enrollment. To determine continuous enrollment for a Medicaid beneficiary for whom enrollment is verified monthly, the member may not have more than a 1-month gap in coverage during each year of continuous enrollment.
Anchor date: December 31 of the measurement year
Benefit: Medical
Event/diagnosis: None

2a1.8 Denominator Exclusions (*Brief narrative description of exclusions from the target population*)**:**

Exclusion: Women who had a bilateral mastectomy or for whom there is evidence of two unilateral mastectomies. Look for evidence of a bilateral mastectomy as far back as possible int he member´s history thorugh Dec 31 of the measurement year.

2a1.9 Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, codes with descriptors, and/or specific data collection items/responses)***:**
Codes to Identify Exclusions
Bilateral mastectomy
CPT: 19180, 19200, 19220, 19240, 19303-19307 WITH Modifier 50 or modifier code 09950*
ICD-9-CM Procedure: 85.42, 85.44, 85.46, 85.48

Unilateral mastectomy (members must have 2 separate occurrences on 2 different dates of service)
CPT: 19180, 19200, 19220, 19240, 19303-19307
ICD-9-CM Procedure: 85.41, 85.43, 85.45, 85.47

*50 and 09950 modifier codes indicate the procedure was bilateral and performed during the same operative session.

2a1.10 Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, codes with descriptors, definitions, and/or specific data collection items/responses)***:**
None

2a1.11 Risk Adjustment Type *(Select type. Provide specifications for risk stratification in 2a1.10 and for statistical model in 2a1.13)***:** No risk adjustment or risk stratification   **2a1.12 If "Other," please describe:**

2a1.13 Statistical Risk Model and Variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development should be addressed in 2b4.)***:**
NA

2a1.14-16 Detailed Risk Model Available at Web page URL (or attachment). Include coefficients, equations, codes with descriptors, definitions, and/or specific data collection items/responses. Attach documents only if they are not available on a webpage and keep attached file to 5 MB or less. NQF strongly prefers you make documents available at a Web page URL**.** Please supply login/password if needed**:**

2a1.17-18. Type of Score: Rate/proportion

2a1.19 Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)***:**
Better quality = Higher score

2a1.20 Calculation Algorithm/Measure Logic*(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)***:**
Step 1. Determine the eligible population. The eligible population is all members who satisfy all specified criteria, including any age, continuous enrollment, benefit, event, or anchor date enrollment requirement.
Step 2. Search administrative systems to identify numerator events for all members in the eligible population.

Step 3. If applicable, for members for whom administrative data do not show a positive numerator event, search administrative data for an exclusion to the service/procedure being measured.
Step 4. Exclude from the eligible population members from step 3 for whom administrative system data identified an exclusion to the service/procedure being measured.
Step 5. Calculate the rate.

2a1.21-23 Calculation Algorithm/Measure Logic Diagram URL or attachment:

2a1.24 Sampling (Survey) Methodology. If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate)**:**
NA

2a1.25 Data Source (*Check all the sources for which the measure is specified and tested*). If other, please describe**:**
 Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record

2a1.26 Data Source/Data Collection Instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.*)**:** Healthcare Effectiveness Data Information Set (HEDIS)

2a1.27-29 Data Source/data Collection Instrument Reference Web Page URL or Attachment:

2a1.30-32 Data Dictionary/Code Table Web Page URL or Attachment:

2a1.33 Level of Analysis (*Check the levels of analysis for which the measure is specified and tested*)**:**
Clinician : Group/Practice, Clinician : Individual, Health Plan, Population : State

2a1.34-35 Care Setting (*Check all the settings for which the measure is specified and tested*)**:** Ambulatory Care : Clinician Office/Clinic

---

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample (*Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*)**:**
HEDIS Health Plan performance data from 2010

2a2.2 Analytic Method (*Describe method of reliability testing & rationale*)**:**
Reliability was estimated by using the beta-binomial model. Beta-binomial is a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® health plan measures. The beta-binomial model assumes the plan score is a binomial random variable conditional on the plan´s true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates. The beta distribution can be symmetric, skewed or even U-shaped.
Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability

in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
Reliability statistic for breast cancer screening:
Commercial 2010: 0.997989
Medicaid 2010: 0.993524
Medicare 2010: 0.991807

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H⦿ M⦿ L◉ I ⦿

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* **are consistent with the evidence cited in support of the measure focus (***criterion 1c***) and identify any differences from the evidence:**
There are some issues with the measure´s current age range due to some differences in the guidelines (as described in the Importance section in this form).

2b2. Validity Testing. *(Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)***:**
NCQA tested the measure for face validity using several panels of stakeholders with specific expertise in measurement and women's health. Panels included representatives from key stakeholder groups, including the American Cancer Society, Health Dialog, family physicians, health plans, state and researchers (See list of members of measure advisory panel for Breast Cancer Screening). Experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area. In addition to the clinical expert panels, NCQA reviews measures with technical panels and offers measures for public comment, and all measures are audited.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)***:**
Through reviews by our clinical expert panel, technical panels, CPM, auditors and public comment, face validity, precision and completeness of administrative data collection has been reviewed and found to be acceptable. In addition, data are reviewed on a yearly basis for accuracy and completeness by NCQA and adjustments are made to specifications as needed based on these reviews. The breast cancer screening measure results correlate well with the cervical cancer screening measure which contributes to the validation and the efficacy of the measure design since both measures are assessing similar preventive health interventions for women.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of*

*measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)***:**

2b3.3 Results (*Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
Breast Cancer Screening is a long-standing HEDIS measure. The measure allows exclusions for a bilateral mastectomy or two unilateral mastectomies, as breast cancer screening is not clinically indicated for women who have had these procedures.

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample (*Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**
NA

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)***:**
NA

2b4.3 Testing Results (*Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)***:**
NA

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: The measure assesses breast cancer in a general population of women; risk adjustment is not indicated. However, measure reporting is stratified by product line in order to prevent inappropriate comparisons across commercial, Medicaid and Medicare health plans.

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample (*Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**
HEDIS Health Plan data

2b5.2 Analytic Method (*Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)***:**
Comparison of means and percentiles; analysis of variance against established benchmarks

2b5.3 Results (*Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance):*
 Commercial Health Plan Rates
Data analysis demonstrate that methods for scoring and analysis of the specified measure allow for identification of statistically significant and meaningful differences in performance.
Results:
Year   2009   2008   2007

```
N            244    253    258
MEAN         71.3   70.2   69.1
STDEV        6.08   6.18   5.94
STDERR       0.39   0.39   0.37
MIN          55.1   47.5   52.2
MAX           90    85.1   84.4
P10          64.2   62.7   61.9
P25          67.0   66.2   64.9
P50          70.7    70    68.5
P75          75.3   74.2   73.5
P90          80.1   78.7   78.2
```

Medicare Health Plans

```
Year        2009    2008    2007
N            290    255    229
MEAN         69.3    68    67.3
STDEV        10.0   11.2   11.7
STDERR       0.59    0.7    0.77
MIN          37.8   13.9    19
MAX          94.7   95.2   90.4
P10          55.7   53.2   52.3
P25          62.5   60.5   60.7
P50          69.8   67.8    68
P75          75.6   75.2   75.1
P90          82.7   82.9   83.5
```

Medicaid Health Plans

```
Year        2009    2008    2007
N            144    137    138
MEAN         52.4   50.8   49.8
STDEV        10.2   10.3   9.36
STDERR       0.85   0.88    0.8
MIN          25.2   18.5   9.38
MAX          78.4   81.7   77.1
P10          39.8   38.6   38.8
P25          46.2    45    44.3
P50           52    50.5    50
P75          59.6   57.4   56.1
P90          63.8    63    61.2
```

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample (*Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*)**:**

2b6.2 Analytic Method (*Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure*)**:**
This measure pulls from administrative claims data.

2b6.3 Testing Results (*Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted*)**:**

2c. Disparities in Care: H⦾ M⦾ L⦾ I ⦾ NA⦾ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results (*Scores by stratified categories/cohorts*)**:** The measure is not stratified to detect disparities.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:
The measure can be stratified by race/ethnicity or other factors. Health plans have demonstrated that they can stratify HEDIS measures if they have complete data on race/ethnicity, for example. In this way, health plans have been able to analyze their rates and develop targeted quality improvement strategies as appropriate. NCQA has a program called the Multicultural Health Care Distinction program that outlines ways in which a plan can collect data to enable disparities analyses and conduct quality improvement activities to reduce health care disparities.

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, **met?** (*Reliability and Validity must be rated moderate or high*) Yes⦾ No⦾
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

3. USABILITY

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. (**evaluation criteria**)

C.1 Intended Actual/Planned Use (*Check all the planned uses for which the measure is intended*)**:** Professional Certification or Recognition Program, Public Reporting, Quality Improvement (Internal to the specific organization), Quality Improvement with Benchmarking (external benchmarking to multiple organizations), Regulatory and Accreditation Programs

3.1 Current Use (*Check all that apply; for any that are checked, provide the specific program information in the following questions*)**:** Public Reporting, Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

3a. Usefulness for Public Reporting: H⦾ M⦾ L⦾ I ⦾
(*The measure is meaningful, understandable and useful for public reporting.*)

3a.1. Use in Public Reporting - disclosure of performance results to the public at large (*If used in a public reporting program, provide name of program(s), locations, Web page URL(s)*). If not publicly reported in a national or community program, state the reason AND plans to achieve public reporting, potential reporting programs or commitments, and timeline, e.g., within 3 years of endorsement: **[For *Maintenance* –** *If not publicly reported, describe progress made toward achieving disclosure of performance results to the public at large and expected date for public reporting; provide rationale why continued endorsement should be considered.***]**
This measure is used in public reporting for plans through Healthcare Effectiveness Data and Information Set (HEDIS) and is reported through venues such as the annual State of Healthcare Quality report, Quality Compass, and America's Best Health Plans. In addition, this measure is included in the proposed Medicaid Adult Core Set.

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

3a.2.Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting. If usefulness was demonstrated (e.g., focus group, cognitive testing), describe the data, method, and results: HEDIS measures adhere to the desirable attributes of scientific acceptability, feasibility and usability. The measures provide performance rates that are audited for consistency and accuracy. NCQA provides benchmarks against which health plans may compare themselves.

3.2 Use for other Accountability Functions (payment, certification, accreditation). If used in a public accountability program, provide name of program(s), locations, Web page URL(s): This is a measure in the HEDIS measure set and is used in NCQA's Health Plan Accreditation program.

3b. Usefulness for Quality Improvement: H⦿ M⦿ L⦿ I ⦿
(*The measure is meaningful, understandable and useful for quality improvement.*)

3b.1. Use in QI. If used in quality improvement program, provide name of program(s), locations, Web page URL(s):
[For _Maintenance_ – *If not used for QI, indicate the reasons and describe progress toward using performance results for improvement*].
This is a measure in the HEDIS measure set and is used in NCQA's Health Plan Accreditation program.

3b.2. Provide rationale for why the measure performance results are meaningful, understandable, and useful for quality improvement. If usefulness was demonstrated *(e.g., QI initiative),* describe the data, method and results:
HEDIS measures adhere to the desirable attributes of scientific acceptability, feasibility and usability. The measures provide performance rates that are audited for consistency and accuracy. NCQA provides benchmarks against which health plans may compare themselves.

Overall, to what extent was the criterion, *Usability*, met? H⦿ M⦿ L⦿ I ⦿
Provide rationale based on specific subcriteria:

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. **(evaluation criteria)**

4a. Data Generated as a Byproduct of Care Processes: H⦿ M⦿ L⦿ I ⦿

4a.1-2 How are the data elements needed to compute measure scores generated? *(Check all that apply).*
Data used in the measure are:
generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

4b. Electronic Sources: H⦿ M⦿ L⦿ I ⦿

4b.1 Are the data elements needed for the measure as specified available electronically *(Elements that are needed to compute measure scores are in defined, computer-readable fields)*: ALL data elements in electronic health records (EHRs)

4b.2 If ALL data elements are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources:

4c. Susceptibility to Inaccuracies, Errors, or Unintended Consequences: H⦿ M⦿ L⦿ I ⦿

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

4c.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measurement identified during testing and/or operational use and strategies to prevent, minimize, or detect. If audited, provide results:
All measures that are used in NCQA programs are audited for accuracy.

4d. Data Collection Strategy/Implementation: H⦿ M⦿ L⦿ I ⦿

A.2 Please check if either of the following apply (*regarding proprietary measures*)**:** Proprietary measure
4d.1 Describe what you have learned/modified 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 (*e.g., fees for use of proprietary measures*)**:**
NCQA assesses the feasibility of the measure through field testing and solicitation of feedback from multi-stakeholder groups, including measure users (i.e. health plans). This measure was determined to be feasible and usable. Because it is a measure that pulls from administrative claims, it can be collected with less burden than measures that require medical record review. In addition, the measure has been specified for EHRs.

Overall, to what extent was the criterion, *Feasibility*, met? H⦿ M⦿ L⦿ I ⦿
Provide rationale based on specific subcriteria:

OVERALL SUITABILITY FOR ENDORSEMENT

Does the measure meet all the NQF criteria for endorsement? Yes⦿ No⦿
Rationale:

If the Committee votes No, STOP.
If the Committee votes Yes, the final recommendation is contingent on comparison to related and competing measures.

---

5. COMPARISON TO RELATED AND 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 before a final recommendation is made.

5.1 If there are related measures (*either same measure focus or target population*) **or competing measures** (*both the same measure focus and same target population*), **list the NQF # and title of all related and/or competing measures:**

5a. Harmonization

5a.1 If this measure has EITHER the same measure focus OR the same target population as NQF-endorsed measure(s): Are the measure specifications completely harmonized?

5a.2 If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden:

5b. Competing Measure(s)

5b.1 If this measure has both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (*e.g., a more valid or efficient way to*

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

measure quality); **OR provide a rationale for the additive value of endorsing an additional measure.** *(Provide analyses when possible)*:

NA - the measures have a different focus and a different target population. Measure 0623 focuses on surveillance for women with a history of breast cancer. NCQA´s measure focuses on secondary prevention in a general population of women.

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner): National Committee for Quality Assurance, 1100 13th Street, NW, District Of Columbia, 20005

Co.2 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Co.3 Measure Developer if different from Measure Steward: National Committee for Quality Assurance, 1100 13th St, NW, Suite 1000, Washington, District Of Columbia, 20005

Co.4 Point of Contact: Jill Marie, Farrell, farrell@ncqa.org, 202-955-3599-

Co.5 Submitter: Rita, Lewis, MPH, lewis@ncqa.org, 202-955-5102-, National Committee for Quality Assurance

Co.6 Additional organizations that sponsored/participated in measure development:

Co.7 Public Contact: Bob, Rehm, Assistant Vice President, Performance Measurement, Rehm@ncqa.org, 202-955-1728-, National Committee for Quality Assurance

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The NCQA Breast Cancer MAP advised NCQA during measure development. They evaluated the way staff specified measures, assessed the content validity of measures, and reviewed field test results. The MAP consisted of a balanced group of experts. Note that, in addition to the MAP, we also vetted these measures 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, in addition to the MAP.

Kathy Coltin, MPH
Lance Lang, MD
Dorothy Mann, PhD
Saralyn Mark, MD Phone: 202-230-4101
Robin Richman, MD, FACOG
Robert Smith, PhD
Eric Tangalos, MD

Ad.2 If adapted, provide title of original measure, NQF # if endorsed, and measure steward. Briefly describe the reasons for adapting the original measure and any work with the original measure steward: NA

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.3 Year the measure was first released: 2000
Ad.4 Month and Year of most recent revision: 2009
Ad.5 What is your frequency for review/update of this measure? Approximately every three years;

earlier or later if warranted by evidence/guidelines
Ad.6 When is the next scheduled review/update for this measure? 07, 2012

Ad.7 Copyright statement: © 2000 by the National Committee for Quality Assurance
1100 13th Street, NW, Suite 1000
Washington, DC 20005

Ad.8 Disclaimers:

Ad.9 Additional Information/Comments:

Date of Submission (*MM/DD/YY*): 07/12/2011