NQF #0037 Osteoporosis testing in older women

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 #: 0037  
NQF Project: Population Health: Prevention Project

(for Endorsement Maintenance Review)
Original Endorsement Date: Aug 10, 2009  Most Recent Endorsement Date: Aug 10, 2009

BRIEF MEASURE INFORMATION

De.1 Measure Title: Osteoporosis testing in older women
Co.1.1 Measure Steward: National Committee for Quality Assurance
De.2 Brief Description of Measure: Percentage of female patients aged 65 and older who reported receiving a bone density test (BMD) to check for osteoporosis

2a1.1 Numerator Statement: The number of patients in the denominator who responded “yes” to the question, “Have you ever had a bone density test to check for osteoporosis, sometimes thought of as “brittle bones”? This test may have been done to your back, hip, wrist, heel, or finger.”

2a1.4 Denominator Statement: Women 65 and older as of December 31 of the measurement year who answered “yes” or “no” to the question, “Have you ever had a bone density test to check for osteoporosis, sometimes thought of as “brittle bones”? This test may have been done to your back, hip, wrist, heel, or finger.”

2a1.8 Denominator Exclusions:

1.1 Measure Type: Process
2a1.25-26 Data Source: Patient Reported Data/Survey

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):

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):

1. IMPACT, OPPORTUNITY, 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**

1. **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.)

2. **Subject/Topic Areas (Check all the areas that apply):**
   - Musculoskeletal: Osteoporosis, Prevention

3. **Cross Cutting Areas (Check all the areas that apply):**
   - Population Health

**1a Demonstrated High Impact Aspect of Healthcare:**

- Affects large numbers
- Patient/societal consequences of poor quality
- Severity of illness

**1a.2 If “Other,” please describe:**

**1a.3 Summary of Evidence of High Impact (Provide epidemiologic or resource use data):**

Osteoporosis is the most common metabolic bone disease and is characterized by low bone mineral density and structural deterioration of bone tissue causing bone fragility and increasing one’s risk of fractures (NIAMS, 2010). It can occur at any age; however one’s risk increases with age. About 44 million Americans live with either osteoporosis or osteopenia (lower than normal bone mineral density that increases risk of osteoporosis), and of this group 68% are women (NIAMS, 2010).

Currently in the US, the estimated national direct expenditures for osteoporosis and related fractures total approximately $14 billion annually (NIAMS, 2010). Since these expenditures do not include indirect costs such as lost productivity or wages, the true financial impact of osteoporosis is extremely underestimated. Experts predict that by 2025 osteoporosis will cost approximately $25.3 billion each year (NOF, 2010). Osteoporotic fractures are responsible for more than 432,000 hospital admissions, almost 2.5 million medical office visits, and about 180,000 nursing home admissions each year (PhysWeeklyArchives.com, March 2, 2009).

Primary osteoporosis often follows menopause in women while medications or other medical conditions and diseases can cause secondary osteoporosis. Many are unaware they have the disease until they break a bone; most commonly breaking a hip, the spine or a wrist. Broken bones are extremely dangerous for older adults. Nearly twenty percent of older adults who suffer a hip fracture will die within a year from complications either related to the break itself or the surgery needed to repair it. Many of those who survivors will never return to pre-fracture functional status which oftentimes forces them to need long-term nursing home care (National Osteoporosis Foundation, 2010). Osteoporosis causes nearly 1.5 million fractures each year including 300,000 hip fractures, 700,000 spinal fractures, 250,000 wrist fractures and over 300,000 other fractures (NIAMS, 2010). Those individuals who have had even a single fracture have a much higher risk of new fractures. Women who have had a history of vertebral fracture are four times more likely to experience a new fracture within the 15-year follow-up (Harvard Health, 2010).

Both the National Osteoporosis Foundation and the US Preventive Services Task Force agree that all women, 65 and older, should be screened routinely with bone mineral density tests. Despite these two group’s strong recommendations and numerous public health campaigns, screening rates are still relatively low. A systematic review of 51 articles examining bone mineral density (BMD) testing trends from 1992 to 2002 found screening frequencies among at-risk patients ranged from 1% to 47%. Medicare claim trends for BMD testing among patients 65 and over increased by nearly 50% from 1999 to 2005, showing 30% of all female Medicare beneficiaries had received at least one BMD test. Although rates have increased over time, there is still room for improvement (Grover et al., 2009).

Burge et al. used modeling to predict the incident osteoporosis-related fractures and subsequent costs in the United States through the year 2025. At that time, annual fractures and costs are projected to rise by almost 50 percent. The most rapid growth is estimated for people 65-74 years of age, with an increase of greater than 87 percent. Furthermore, an increase of nearly 175 percent is projected for subpopulations such as Hispanics, African Americans and men (Burge, et al., 2007). Despite having a lower fracture risk, older men tend to have worse outcomes after fracture and poorer treatment rates (Cawthon, 2011). One recent study found an increase in treatment rates for men and low treatment variability between race/ethnic groups in a healthcare system using electronic medical records. The electronic medical records helped identify care gaps and gave continued reminders to providers until the care gaps were closed (Navarro, et al., 2011). Another study developed a fracture liaison service (FLS) and was able to obtain a high level of persistence with osteoporosis treatment. The authors claimed that since follow-up and treatment renewal were under routine daily practice, these results underscore the importance of initial prescription conditions and highlights an interest in medical networks such as the FLS (Boudou, et al., 2011).
A significant risk has been reported in people of all ethnic backgrounds. Non-Hispanic Caucasian and Asian women aged 50 and older are at particular risk for osteoporosis and low bone mass. While, twenty percent of non-Hispanic Caucasian women aged 50 or older are estimated to have osteoporosis, women from other racial/ethnic backgrounds are also at risk for osteoporosis and low bone mass. Five percent of non-Hispanic black women over age 50 are estimated to have osteoporosis; an estimated additional 35% have low bone mass that puts them at risk of developing osteoporosis. Ten percent of Hispanic women aged 50 and older are estimated to have osteoporosis; an estimated additional 49% are estimated to have low bone mass. When compared with other ethnic/racial groups, risk is increasing most rapidly among Hispanic women. Osteoporosis is considered to be under recognized and under-treated in both Caucasian and African American women (National Osteoporosis Foundation, 2010).

There is a misconception that osteoporosis is only a concern for white women, which is delaying the prevention and treatment in other ethnic populations of women who currently believe they are not at risk for the disease. African-American and Hispanic women are less likely to believe they are at risk for osteoporosis and feel osteoporosis is not a major health concern as some other diseases (NIAMS, 2010). Prevention efforts should target all women, irrespective of their race/ethnicity, especially if they have multiple risk factors (Cauley, 2011).

1a.4 Citations for Evidence of High Impact cited in 1a.3:

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

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:
The benefits of screening for osteoporosis include the detection of lower bone density mass and the prevention of fractures, particularly in older women. The United States Preventive Task Force (USPSTF) found good evidence that the risk of osteoporosis and fractures increases with age and other factors, that bone density measurements accurately predict the risk of fractures in the short-term. The USPSTF found that there are at least moderate benefits of screening for women at increased risk by virtue of age, and recommends women aged 65 and older be screened routinely for osteoporosis (USPSTF 2010).


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.]
Osteoporosis testing in older women Rate - Total Data Element; 2009; 2008; 2007;
These performance reports indicate room for improvement.

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]

The data are performance rates from all health plans participating in the HEDIS measure set. There were 1234 plan submissions for this measure. NCQA collects data directly from Health Plan Organizations and Preferred Provider Organizations via a data submission portal - the Interactive Data Submission System (IDSS). NCQA assigns a sub-ID by an accreditable identity based on the legal entity and management structure that supports the product lines/products that NCQA accredits. Each accreditation is legally accountable entity provides to members and representation of an organization and delivery structure that is meaningful to members.

1b.4 Summary of Data on Disparities by Population Group: [For Maintenance – Descriptive statistics for performance results for this measure by population group]

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]

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.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Quality</th>
<th>Consistency</th>
<th>Does the measure pass subcriterion 1c?</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-H</td>
<td>M-H</td>
<td>M-H</td>
<td>Yes □</td>
</tr>
<tr>
<td>L</td>
<td>M-H</td>
<td>M</td>
<td>Yes □ IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise No □</td>
</tr>
<tr>
<td>M-H</td>
<td>L</td>
<td>M-H</td>
<td>Yes □ IF potential benefits to patients clearly outweigh potential harms: otherwise No □</td>
</tr>
<tr>
<td>L-M-H</td>
<td>L-M-H</td>
<td>L</td>
<td>No □</td>
</tr>
</tbody>
</table>

Health outcome – rationale supports relationship to at least one healthcare structure, process, intervention, or service

Does the measure pass subcriterion 1c? 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):

As osteoporotic fractures are a major health issue for many older women, this measure seeks to ensure that appropriate, recommended testing is provided.
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):
The United States Preventive Services Task Force recommends that women age 65 and older receive routine screening for osteoporosis (USPSTF 2011). There are numerous advanced screening methods for osteoporosis, yet the rate for which postmenopausal women are receiving these screenings is very low. The Dual-energy x-ray absorptiometry (DEXA) is considered the “gold standard” bone density test used in screening for osteoporosis as it quantitatively calculates the photon absorption of the minerals in bone tissue. However, in a 2005 Medicare claims survey asking women 65 and older if they had been given a diagnostic bone mineral density (BMD) exam in the last year, only 12.9% reported they had (Sego, 2010).

Studies show the prevention of fractures and falls have an effect on the quality of life and physical functioning of elderly people. Osteoporotic patients with vertebral fractures had worse scores for domains of physical function and social function general health perception. Vertebral fractures and a low femoral BMD impair QOL perception (Romagnoli 2004).

According to a study conducted by Kaiser Permanente, consistent management of patients at risk for osteoporosis could reduce the rate of reported hip fractures in the United States by 25% (ScienceDaily, 2009).

1c.5 Quantity of Studies in the Body of Evidence (Total number of studies, not articles):

1c.6 Quality of Body of Evidence (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):

1c.7 Consistency of Results across Studies (Summarize the consistency of the magnitude and direction of the effect):

1c.8 Net Benefit (Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms):

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: US Preventive Services Task Force (USPSTF). The USPSTF recommends screening for osteoporosis in women aged 65 years or older and in younger women whose fracture risk is equal to or greater than that of a 65-year-old white woman who has no additional risk factors. This is a B recommendation.

The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of screening for osteoporosis in men. This is an I statement.

National Osteoporosis Foundation (NOF): In women age 65 and older and men age 70 and older, recommend bone mineral density (BMD) testing. In postmenopausal women and men age 50-69, recommend BMD testing when you have concern based on their risk factor profile.

American Association of Clinical Endocrinologists (AACE) recommends women age 65 years and older (Grade B, Best Evidence Level 2) and all younger postmenopausal women at increased risk of fracture be screened for osteoporosis (Grade C, Best Evidence Level 2).

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:  B

1c.14 Summary of Controversy/Contradictory Evidence:  While the majority of individuals affected by osteoporosis are women, current studies are exploring the benefits of osteoporosis screening in males. Seven percent of non-Hispanic Caucasian and Asian men age 50 or older are estimated to have osteoporosis, while 35% are estimated to have low bone mass. Four percent of non-Hispanic black men age 50 and older are estimated to have osteoporosis, while 19% are estimated to have low bone mass. Three percent of Hispanic males age 50 or older are estimated to have osteoporosis, while 23% are estimated to have low bone mass (National Osteoporosis Foundation, 2010). Other strong predictors for increased risk of osteoporosis include age, low body weight, physical inactivity, and weight loss (Shekell 2007). Currently there only a limited number of studies publish which identify osteoporosis screening tools in men. A recent cost-effectiveness analysis using Markov modeling indicated universal DXA screening in men would not be cost effective (Schousboe 2006). Six out of 10 males have osteoporosis by age 65 years and early screening based on risk factors could prevent osteoporosis-related fractures. Currently, the American College of Physicians (ACP) recommends bone thickness measurement with DXA for men who have risk factors for osteoporosis and who are willing and able to take drugs (Qaseem 2008).

1c.15 Citations for Evidence other than Guidelines (Guidelines addressed below):
**1c.16 Quote verbatim, the specific guideline recommendation (including guideline # and/or page #):**

US Preventive Services Task Force (USPSTF): The USPSTF recommends screening for osteoporosis in women aged 65 years or older and in younger women whose fracture risk is equal to or greater than that of a 65-year-old white woman who has no additional risk factors. This is a B recommendation. The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of screening for osteoporosis in men. This is an I statement.

National Osteoporosis Foundation (NOF): In women age 65 and older and men age 70 and older, recommend bone mineral density (BMD) testing. In postmenopausal women and men age 50-69, recommend BMD testing when you have concern based on their risk factor profile.

American Association of Clinical Endocrinologists (AACE) recommends women age 65 years and older (Grade B, Best Evidence Level 2) and all younger postmenopausal women at increased risk of fracture be screened for osteoporosis (Grade C, Best Evidence Level 2).


**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:** A—Strongly Recommended: The USPSTF recommends the service. There is high certainty that the net benefit is substantial.. B—Recommended: The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial., C—No Recommendation: The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small.

**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:**

**1c.24 Rationale for Using this Guideline Over Others:** It is NCQA policy to use guidelines that are evidence-based, applicable to physicians and other healthcare providers, and developed by a national specialty organization or government agency.

NCQA convened an expert panel of diverse stakeholders to review the guidelines and evidence for this measure. The panel determined the measure was scientifically sound using the full body of evidence and guidelines for this measure concept.
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: High |

Was the threshold criterion, Importance to Measure and Report, met?

(1a & 1b must be rated moderate or high and 1c yes) Yes [ ] No [x]

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

S.2 If yes, provide web page URL:

2a. RELIABILITY. Precise Specifications and Reliability Testing: H [ ] M [ ] L [x] 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):

The number of patients in the denominator who responded “yes” to the question, “Have you ever had a bone density test to check for osteoporosis, sometimes thought of as “brittle bones”? This test may have been done to your back, hip, wrist, heel, or finger.”

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

Measurement year (one calendar year)

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:

Responses of "yes" to Q52 in the Medicare Health Outcomes Survey (HOS)

“Have you ever had a bone density test to check for osteoporosis, sometimes thought of as "brittle bones"? This test may have been done to your back, hip, wrist, heel, or finger.”

2a1.4 Denominator Statement (Brief, narrative description of the target population being measured):

Women 65 and older as of December 31 of the measurement year who answered “yes” or “no” to the question, “Have you ever had a bone density test to check for osteoporosis, sometimes thought of as “brittle bones”? This test may have been done to your back, hip, wrist, heel, or finger.”

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

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

Measurement year (one calendar 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):
Female Medicare members age 65 and above

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

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):

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):

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.):

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.):
Results are calculated by NCQA using data collected in the combined HOS Baseline and Follow-Up Survey samples from the same measurement year.

The Medicare Advantage Organization (MAO) must achieve a denominator of at least 100 to obtain a reportable result. If the denominator is less than 100, NCQA assigns a measure result of Not Applicable (N/A).

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):
The measure is collected in the Medicare Health Outcomes Survey (HOS). Medicare Advantage Organizations (MAOs) reporting the measure must contract with a NCQA-Certified HOS Survey Vendor to administer the survey. A minimum of 1,200 members per MAO are randomly selected for the survey. Plan-level results are calculated by NCQA using data collected in the combined HOS Baseline and Follow-Up Survey samples from the same measurement year. MAOs must achieve a denominator of at least 100 to
obtain a reportable result. If the denominator is less than 100, NCQA assigns a measure result of NA.
NCQA outlines the sampling criteria for all HOS measures. The complete data collection method and sampling guidelines are outlined in NCQA’s HEDIS Technical Specifications for the HOS, Volume 6.

2a1.25 Data Source (Check all the sources for which the measure is specified and tested). If other, please describe:
Patient Reported Data/Survey

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.): Medicare Health Outcomes Survey

2a1.27-29 Data Source/data Collection Instrument Reference Web Page URL or Attachment: URL http://www.hosonline.org/Content/SurveyInstruments.aspx

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, Clinician : Team, Health Plan, Integrated Delivery System, Population : National

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

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 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 for this measure was calculated as 0.949356.

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:
They are consistent.

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 a panel of stakeholders with specific expertise in measurement and geriatric care. This panel included representatives from key stakeholder groups, including geriatricians, health plans, Medicare officials, and researchers. Experts 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.

The survey question comprising this measure was tested in a representative sample of elderly respondents between January and March 2005, to examine validity of concepts and comprehension of survey questions so as to elicit a valid response.

This measure captures women who have received a test for osteoporosis, regardless of whether or not the purpose of the test was to screen, confirm a diagnosis or monitor the treatment of a patient with osteoporosis. While the measure does not correlate exclusively with the USPSTF guidelines and does not specify how many women under 65 are being screened, it does capture a meaningful aspect of care. This measure correlates with other HEDIS measures on the Medicare HOS that seek to improve diagnosis of conditions such as the rate of discussing urinary incontinence. The Falls Risk Management measure also complements the osteoporosis testing measure, as many patients with undiagnosed osteoporosis are likely to experience a fracture after a fall is recommended for bone health, and improved screening, diagnosis and treatment of osteoporosis will lead to improved falls prevention.

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):
This measure was deemed valid by the expert panel.

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):

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):

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

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):
### 2b.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:

This process measure assesses osteoporosis treatment in older women for a general population; risk adjustment is not indicated.

### 2b.5. 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:*

Data analysis demonstrates that methods for analyzing the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.

#### 2b5.2 Analytic Method

*Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance:*

Comparison of means and percentiles; MAOs with enrollment size <100 are not used because the small sample would not allow for statistically significant differences in performance.

#### 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:*

### 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:*

Since the information for this measure comes from one data source using the same survey question and survey, data comparability is not an issue.

#### 2b6.2 Analytic Method

*Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure:*

#### 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:*

This 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:

### 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 Purpose/Use (Check all the purposes and/or uses for which the measure is intended): Payment 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, Payment Program

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 only through Healthcare Effectiveness Data and Information Set (HEDIS) and is reported through venues such as the annual State of Healthcare Quality report, Quality Compass, America’s Best Health Plans.

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:

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):

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 measure is a measure in the Healthcare Effectiveness Data and Information Set (HEDIS).

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:

Not applicable. NCQA reports on performance of health plans and providers nationally. Our results are not part of an internal NCQA QI program.

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:
Other
Patient-reported health survey
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): No data elements are in electronic sources

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: Data are collected via a patient mail (or telephone) survey. Electronic surveys may be available in the future.

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

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: Susceptibility to inaccuracies, errors, or unintended consequences and the ability to audit the data items to detect such problems are identified. All measures that are used in NCQA programs are audited.

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

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): This measure is precisely specified using the survey data collection method. This measure has detailed, precise specifications that clearly define the numerator, denominator, data sources, allowable values, methods of measurement and method of reporting.

The Osteoporosis Technical Subgroup originally considered this as an administrative claims measure, assessing the number of women over 65 who had a BMD to screen for osteoporosis. However, a measure that only captures BMDs for screening purposes only, but not for diagnosing or monitoring osteoporosis, requires accurately excluding women who either had a previous diagnosis of osteoporosis, treatment for osteoporosis, or was previously screened regardless of the length of time, using administrative claims data. Given the limitations of administrative data, an administrative measure would not be most feasible. Moreover, the guidelines does not specify how often women should be screened making an administrative measure challenging without undue burden for data collection. Therefore a survey measure assessing whether all women over 65 have at least had one bone density test was considered to be the most reliable and accurate method to identify whether an elderly woman has ever received an appropriate test for osteoporosis for either screening or monitoring purposes. Excluding women with a diagnosis or treatment of osteoporosis was therefore also not necessary.

To ensure survey questions are reliable and valid, the question was cognitively tested to identify any difficulty with the wording and response choices. Results of cognitive testing were used to further refine the survey questions comprising the measure.

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...
NQF #0037 Osteoporosis testing in older women

measure focus and same target population), list the NQF # and title of all related and/or competing measures:
0035 : Fall risk management in older adults: a. Discussing fall risk, b. Managing fall risk

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? Yes

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

CONTACT INFORMATION


Co.2 Point of Contact: Bob, Rehm, Assistant Vice President, Performance Measurement, Rehm@ncqa.org, 202-955-1728-

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

Co.4 Point of Contact: Bob, Rehm, Assistant Vice President, Performance Measurement, Rehm@ncqa.org, 202-955-1728-

Co.5 Submitter: Dawn, Alayon, MPH, CPH, Senior Health Care Analyst, alayon@ncqa.org, 202-955-3533-, 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.

GMAP
Wade Aubry, BCBS Association
Arlene Bieman, University of Toronto and St. Michael's Hospital
Joyce Dubow, AARPPeter Hollmann, BCBS of Rhode Island
Jerry Johnson, University of Pennsylvania
David Martin, Ovations
Cheryl Phillips, On Lok Lifeways
Steven Phillips, Sierra Health Services, Inc.
Scott Sarran, BCBS of Illinois
Eric G Tangalos, Mayo Clinic
Joan Weiss, Health Resources and Services Administration
Neil Wenger, UCLA Division of General Internal Medicine and RAND
The NCQA Geriatric Measurement Advisory Panel advised NCQA during measure development. They evaluated the way staff specified measures, assessed the content validity of measures, and reviewed cognitive testing results. As you can see from the list, the MAP consisted of a balanced group of experts, including representatives from medical research and education, health plans, the federal Medicare program, and older adult associations. 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.

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:

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.3 Year the measure was first released: 2005
Ad.4 Month and Year of most recent revision: 11, 2009
Ad.5 What is your frequency for review/update of this measure? Approximately every three years; sooner if the clinical guidelines have changed significantly.
Ad.6 When is the next scheduled review/update for this measure? 07, 2013

Ad.7 Copyright statement: © June 29, 2011 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