Quality Payment PROGRAM

~ eCQM Example ~

Peer Reviewed Journal Article Requirement

Section 101(c)(1) of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) requires submission of new measures for publication in applicable specialty-appropriate, peer-reviewed journals prior to implementing in the Merit-based Incentive Payment System (MIPS). These measures will be submitted to journal(s) before including any new measure in the final list of annual clinical quality measures (CQM) under MIPS. The measure owner shall provide the required information for article submission under the MACRA per Center for Medicare & Medicaid Services' (CMS) "Call for Measures" submission process.

Measure owners submitting measures into JIRA must complete the required information by the Call for Measures deadline. Some of the information requested below may be listed in specific fields in the JIRA tool; however, to ensure that CMS has all of the necessary information and to avoid delays in the evaluation of your submission, please fully complete this form as an attached Word document in JIRA. The information in JIRA must be consistent with the information below. This includes, but is not limited to:

***Measure Title:* Appropriate Use of DXA Scans in Women Under 65 Years Who Do Not Meet the Risk Factor Profile for Osteoporotic Fracture**

***Domain:* Making care more affordable**

Measure Owner: *Centers for Medicare & Medicaid Services*
Measure Developer: *National Committee for Quality Assurance*
Description: Percentage of female patients aged 50 to 64 without select risk factors for osteoporotic fracture who received an order for a dual-energy x-ray absorptiometry (DXA) scan during the measurement period.

✓ Please choose the appropriate domain from the following list:

- Effective Clinical Care
- Communication and Care Coordination
- Patient Safety
- Person and Caregiver-Centered Experience and Outcomes
- Efficiency and Cost Reduction
- Community/ Population Health

✓ Please refer to resources such as the CMS Needs and Priorities Document posted on the 2018 CMS Pre-Rulemaking website for more information regarding these domains.

I. Statement

- *Background (Why is this measure important?)* This measure encourages the appropriate use of DXA scans in women under 65 years. The 2011 U.S. Preventive Services Task Force (USPSTF) osteoporosis screening guidelines recommend DXA scans in women under age 65 only if they are at risk for osteoporotic fractures (USPSTF 2011). In 2012, the American Academy of Family Physicians (AAFP) released a report citing DXA scans as one of five tests widely considered to be overused by family physicians (AAFP 2012). In

addition, an ad hoc work group of internists from across the nation, convened by the American College of Physicians, identified "performing DXA screening for osteoporosis in women younger than 65 in the absence of risk factors" as one of the 37 commonly used screening or diagnostic tests that do not reflect high-value care (Qaseem et al. 2012). Although there is evidence to support the cost-effectiveness of DXA screening in women older than 65, there is not enough evidence to support screening women younger than 65 who do not meet a risk-factor profile (Lim et al. 2009).

References

AAFP (American Academy of Family Physicians). "American Academy of Fmily Physicians Releases 'Top 5' List Of Possibly Overused Tests and Procedures." Available at https://www.aafp.org/media-center/releases-statements/all/2012/choosingwisely.html. Accessed June 10, 2017.

Lim, L. S., L. J. Hoeksema, and K. Sherin. "Screening for Osteoporosis in the Adult U.S. Population: ACPM Position Statement on Preventive Practice." *American Journal of Preventive Medicine*, vol. 36, no. 4, April 2009, pp. 366–375.

Qaseem, A., P. Alguire, P. Dallas, L. Feinberg, F. Fitzgerald, C. Horwitch, L. Humphrey, et al. "Appropriate Use of Screening and Diagnostic Tests to Foster High-Value, Cost-Conscious Care." *Annals of Internal Medicine*, vol. 156, no. 2, January 2012, pp. 147–149.

USPSTF (U.S. Preventive Services Task Force). "Screening for Osteoporosis: U.S. Preventive Services Task Force Recommendation Statement." *Annals of Internal Medicine,* vol. 154, no. 5, March 2011, pp. 356–364.

✓ Please review for competing measures. If a **similar measure** is identified, provide a rationale to justify the selection of the submitted measure over the current measure.

- *Environmental Scan (Are there existing measures in this area?)* There is one related measure (NQF 0046: Screening or Therapy for Osteoporosis for Women Aged 65 Years and Older); however, this measure encourages DXA scans in older women, which is supported by current clinical guidelines. The proposed measure aims to curb potential overuse in younger women (namely, those ages 50 to 64) without risk factors that would justify such a procedure, and would serve as a counterbalance to the existing measure.

✓ In this section, explain the gap between actual healthcare and ideal healthcare and how this measure will assist in closing the gap. Provide statistical data supporting the **existence of a gap** in healthcare which may include average performance rates, ratios, and performance range.

II. Gap Analysis

- *Provide Evidence for the Measure (What are the gaps and opportunities to improve care?)* The number of women without osteoporosis or risk factors for osteoporotic fracture is estimated to exceed 26 million (Wright et al. 2016). The scientific literature and our measure testing suggest that anywhere from *0.25 to 50 percent of women receive potentially inappropriate DXA screenings*. A recent longitudinal cohort study found that in one regional

health system over a seven-year period, *almost 50 percent of women under 65 without osteoporosis risk factors received a DXA screening* (Amarnath et al. 2015). Testing of this measure using electronic health record (EHR) data for women between the ages of 50 and 65 who received care in two physician practices (denominators: 66,778 and 2,508,693 patients, respectively) found that *rates of potentially inappropriate DXA scans varied from 0.25 to 8.19 percent*. An analysis of claims data for Medicare, Medicaid, and commercial plans found that *68.5 percent of DXA scans performed on women under 65 occurred in women ages 50 to 64*, and *6 to 7 percent of women in this 50- to 64-year-old age group received an "inappropriate" DXA scan*. Although the rate of inappropriate DXA screenings varies significantly across these sources, all indicate a significant quality gap and a need for attention to this area.

References

Amarnath, A. L. D., P. Franks, J. Robbins, G. Xing, and J. J. Fenton. "Underuse and Overuse of Osteoporosis Screening in a Regional Health System: A Retrospective Cohort Study." Journal of General Internal Medicine, vol. 30, no. 12, December 2015, pp. 1733–1740.

Wright, N. C., A. C. Looker, K. G. Saag, J. R. Curtis, E. S. Delzell, S. Randall, and B. Dawson-Hughes. "The Recent Prevalence of Osteoporosis and Low Bone Mass in the United States Based on Bone Mineral Density at the Femoral Neck or Lumbar Spine." Journal of Bone and Mineral Research, vol. 29, no. 11, October 2014, pp. 2520–2526.

- *Expected Outcome (Patient care/patient health improvements, cost savings)*
  Our review of the literature and development of a cost-savings model indicate that improvement in performance of this measure might lower the cost to society by reducing (1) potentially inappropriate scans for those women under age 65 who do not meet the risk-factor profile and (2) long-term health problems due to unnecessary exposure to DXA radiation.

  Reducing the overuse of DXA among U.S. women under 65 who are not at risk for osteoporosis would lower health care expenditures related to DXA imaging. In 2011, national payment rates for office-based DXA testing were estimated to be $97.51 (King and Fiorentino 2011). Using a simple model based on the current population of U.S. women ages 50 to 65, and assuming an 8 percent performance rate as seen in preliminary testing of this measure concept, we estimated that a 10 percent reduction in potentially inappropriate DXA scans for women without risk factors for osteoporosis or osteoporotic fracture could save more than $20 million a year in health care expenditures, taking into consideration only those billing costs directly related to DXA scans.

(The 10 percent decrease would generate an average performance rate of 7.2 percent, the lower score indicating better quality.)

A reduction in unnecessary DXA scans for women without risk of osteoporosis or osteoporotic fracture would also decrease unnecessary exposure to radiation. As with other X-ray-based imaging, the radiation dose from DXA must be kept as low as reasonably possible to limit exposure and the concomitant risk of carcinogenesis (Damilakis et al. 2010). Radiation from DXA scans is generally lower than for other imaging modalities. For example, a typical dose for central DXA is 0.0042 mSv; for X-ray mammography, 0.4 mSv; and for adult abdominal computed tomography, 8.0 mSv (Damilakis et al. 2010). Although radiation exposure is low for DXA, its inappropriate use among women under 65 without risk factors for osteoporosis or osteoporotic fracture exposes these women to radiation that is unnecessary and potentially harmful over the course of a lifetime.

References

Damilakis, J., J. E. Adams, G. Guglielmi, and T. Link. "Radiation Exposure in X-ray-based imaging Techniques in Osteoporosis." *European Radiology*, vol. 10, no. 11, November 2010, pp. 2707–2714.

King, A. B., and D. M. Fiorentino. "Medicare Payment Cuts for Osteoporosis Testing Reduced Use Despite Tests' Benefit in Reducing Fractures." *Health Affairs*, vol. 30, no. 12, December 2011, pp. 2362–2370.

- *Recommendation for the Measure (Is it based on a study, consensus opinion, USPSTF recommendation etc.?)* This measure is based on the USPSTF's osteoporosis guidelines, which recommend screening postmenopausal women younger than 65 for osteoporosis only if they meet a risk-factor profile. The risks for those under 65 who merit osteoporosis screening include, but are not limited to, previous osteoporotic fracture, osteoporosis, rheumatoid arthritis and other conditions associated with secondary osteoporosis, parental history of fractures, body mass index (BMI) less than 21 kg/m^2, long-term use of glucocorticoids, current smoking, or excessive alcohol intake (USPSTF 2011).

✓ This section should list the recommendations that support the measure. Quality measures should reflect **current** guidelines.

Reference

USPSTF (U.S. Preventive Services Task Force). "Screening for Osteoporosis: U.S. Preventive Services Task Force Recommendation Statement." *Annals of Internal Medicine*, vol. 154, no. 5, March 2011, pp. 356–364

III. Reliability/Validity (If applicable)

- *What testing has been performed at the clinician level? Please provide testing results including the N value, Bonnie test case results, correlation coefficient and any other pertinent information or values to be considered*. Yes, we tested the measure's reliability and validity using EHR data from three physician practices. We also examined the frequency of exclusions using EHR data from these three sites and a proprietary database containing claims for patients enrolled in a mix of commercial, Medicaid, and Medicare plans. The EHR data contained information from 1,250 physicians and 215,734 women ages 18 to 64. The claims database contained information on 7,557,333 women ages 18 to 64. (The current measure focuses on women ages 50 to 64, but early analyses considered a broader age range. After we spoke with the experts, we decided to focus on women ages 50 to 64, as they were much more likely to be screened than women younger than 50. We have provided results that conform with the specified age range wherever possible.)

Reliability testing: We used a test-retest approach with bootstrapping, also referred to as a random split-half correlational approach, to estimate the reliability of the measure at the provider level. Patients for each provider were randomly split into two groups. For each group, we computed provider performance scores. We then estimated the Pearson correlation coefficient using the provider performance scores for each group. This process was repeated 2,500 times to improve the precision of our reliability estimates by reducing the variance of the correlation coefficient. We estimated the average reliability coefficient using different denominator thresholds to examine the relationship between denominator size (at the provider level) and estimates of reliability. Measures with reliability coefficients of 0.70 are generally considered adequately reliable (Nunnally and Bernstein 1994). The average reliability coefficient for primary care physicians (PCPs) with 20 or more patients in the denominator was 0.827 (Table 1), which supports its use in a physician-reporting program.

Table 1. Provider reliability scores, by number of patients in the provider's denominator (EHR data)

Provider type	Minimum patients per provider in denominator	Provider *N*	Average reliability coefficient
PCPs	1	269	0.245
PCPs	10	170	0.682

✓ The measure is required to be **fully tested** at the clinician level in order to progress through the MUC process for MAP consideration.

✓ Measure developers are required to submit **Bonnie test cases** to ensure the measure algorithm performs as intended. When measure logic is revised, additional Bonnie testing should be performed.

✓ In the reliability section, please include the N, correlation coefficient, and any other pertinent information or values in a **table format** if possible so that the information can be readily available and easily inferred to support the reliability of the quality measure.

PCPs	20	138	0.827

Notes: Rates were calculated using EHR extracts from sites 1 and 3. Patient counts and reliability analysis are restricted to patients ages 50 to 64, which is the current focus of the measure.

Validity testing, clinician sites. At each of the three sites, we first extracted EHR data for women ages 18 to 64 who had an encounter with a PCP during the measurement period. We then completed chart abstraction on a random sample of 200 patients at the three testing sites. Patients were oversampled on DXA orders (numerator condition). We determined the number of patients to sample through a power analysis designed to achieve 80 percent power to detect differences between the EHR extract and the manually abstracted data. Agreement was assessed between EHR extracts and manually abstracted data at the three sites. Kappa scores greater than 0.75 indicate excellent agreement, whereas scores below 0.40 indicate poor agreement (Fleiss 1981). Two of the testing sites used patient encounter data from 2013, and the third site used encounter data from 2012.

✓ Provide statistical data supporting the validity of the measure in a table format along with the level testing performed.

We used a matched EHR extract and a manual abstract of chart data for a sample of patients to compute kappa agreement statistics to estimate the chance-adjusted agreement between the two data sources for a randomly selected set of patients at each of our testing sites. This approach enabled us to assess the validity of the EHR extract compared with a definitive record of a patient's care. We also reported overall agreement and sensitivity and specificity, using the chart-abstracted data as the gold standard for comparison. Validity data were stratified by site to understand how site characteristics (for example, documentation patterns) affected measure element validity. Chance-adjusted agreement between sites' EHR extracts and manually abstracted data for the numerator condition (DXA order) was excellent at two of the three sites in our testing protocol (Table 2). The third site had agreement equal to chance for DXA orders. The low rate of chance-adjusted agreement was attributable to a lack of EHR documentation for DXA scans (0.5 percent in EHRs versus 48.5 percent in the manually abstracted data). During our site visit, the two sites with high kappa agreement noted that they routinely used a structured field for DXA orders.

Table 2. DXA numerator condition: DXA order (EHR data)

Agreement statistic	Site 1	Site 2	Site 3
Kappa	–0.010	0.915	0.928
Overall	51.0%	96.2%	99.0%
Sensitivity	0.0%	98.5%	100.0%
Specificity	99.0%	95.3%	98.0%

Exclusion frequency: We explored the prevalence of exclusions using EHR data from three sites for women ages 50 to 64 and claims data for women ages 18 to 64 who had a claim for a DXA scan. The results are shown in Table 3. Overall, prevalence rates for most exclusions were typically under 5 percent, with the exceptions of osteoporosis and osteopenia. Although many exclusions have low prevalence, they are based on evidence and add to the face validity of the measure, so we maintained them in the measure. Prevalence of exclusions in claims was typically higher than the EHR data, due in part to the lower prevalence of data elements when compared with other, more definitive data sources (such as chart review). However, several exclusions were not available in claims data, such as BMI, smoking status, and alcohol consumption. Because of the low prevalence of exclusions for this measure, it was not possible to compute reliable agreement statistics for most exclusions. The most prevalent exclusion across the testing sites was current smoker status. At Sites 2 and 3, kappa agreement for this exclusion was very good (0.821 and 0.970, respectively). Table 2 lists the kappa agreement for the exclusions where kappa agreement could be estimated; at Sites 2 and 3, kappa agreement for exclusions ranged from poor to very good, whereas for Site 1, kappa values did not exceed 0.120.

Table 3. Prevalence of exclusions (EHR and claims data)

Exclusion	• Claims (women ages 18–64 with DXA order)	EHR, Site 1 (women ages 50–64)	EHR, Site 2 (women ages 50–64)	EHR, Site 3 (women ages 50–64)
Osteopenia	46.9%	4.3%	57.2%	11.0%
Osteoporosis	26.2%	2.7%	0.1%	0.5%
Chronic liver disease	4.9%	1.9%	15.0%	1.6%
Malabsorption syndromes	3.9%	2.6%	8.1%	1.7%
Hyper-thyroidism	2.5%	1.1%	3.3%	#
Rheumatoid arthritis	3.3%	1.0%	6.9%	1.1%
Type I diabetes	1.5%	0.6%	7.0%	0.5%
Lupus	1.3%	0.7%	3.4%	#
Chronic malnutrition	0.5%	#	1.5%	#
Prior osteoporotic fracture	1.3%	#	--	#
Glucocorticoids (oral only)	0.8%	--	--	--
Hyperpara-thyroidism	1.3%	0.7%	--	#
Psoriatic arthritis	0.3%	#	1.1%	#
Premature menopause	0.9%	--	--	--

End-stage renal disease (ESRD)	0.2%	1.1%	1.0%	1.2%
Immuno-suppressants	0.3%	--	--	--
Ankylosing spondylitis	0.1%	#	#	#
Gastric bypass surgery	0.1%	#	--	#
Cushing's syndrome	0.1%	#	#	#
Ehlers-Danlos syndrome	0.0%	#	#	#
Marfan's syndrome	0.0%	#	#	#
Osteogenesis imperfecta	0.0%	#	#	#
Low BMI	--	16.4%	*	9.8%
Current smoker	--	7.5%	*	12.0%
High alcohol consumption	--	5.5%	*	--

Prevalence based on EHR data was < 0.5%.
* Data exist in EHR but they were not extracted properly for testing.
-- Data not available.

References

Nunnally, J. C., and Bernstein, I. H. *Psychometric Theory*, 3rd ed. New York: McGraw-Hill, 1994.

Fleiss, J. L. *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons, 1981.

- *What were the minimum sample sizes used for reliability results?* As shown in Table 1 above, we assessed provider-level reliability with three minimum sample sizes: 1 patient, 10 patients, and 20 patients. Based on the accepted threshold of 0.70 for reliability coefficients, we recommend establishing a minimum sample size of 20 patients for this measure.

a. Other Information
   - *Is it risk adjusted? If so, how?* This measure is not risk adjusted. *What benchmarking information is available?* We have not studied or established any benchmarks for this measure

IV. Endorsement

- *Provide NQF endorsement status (and ID) and/or other endorsing body.* This measure is not currently NQF endorsed.

✓ Endorsement is not required but it is encouraged.

V. Summary

- *Alignment with CMS Meaningful Measures Initiative or MACRA (If applicable).* This measure is consistent with the priority in MACRA for "measures of appropriate use of services, including measures of over use" (MACRA 2015).
- *Importance to MIPS or other CMS programs*.This measure promotes evidence-based care by discouraging overuse of a common procedure. It complements an existing measure of appropriate use.
- *Rationale: Use of measure for inclusion in program (specialty society, regional collaborative, other)*.This measure has not yet been implemented into an existing program; however, we believe this measure would receive support as a meaningful and useful quality-care concept. It is aligned with the National Quality Strategy, CMS's Quality Measure Development Plan, the Choosing Wisely™ Campaign, and ongoing concerns among clinicians regarding the overuse of diagnostic procedures. Our testing results and the feedback we have received from expert panels indicate that the measure can be successfully implemented to assess physicians' performance.
- *Public reporting (if applicable)* As this measure has not yet been implemented, it is not yet publicly reported.
- Preferable relevant Peer-Review Journal for publication. We recommend submitting this measure to a journal related to either preventive medicine or primary care (e.g., the *American Journal of Preventive Medicine* or the *Journal of Family Medicine and Primary Care*) or women's health (e.g., the *Journal of Women's Health or Women's Health Issues*).

Reference

MACRA. "Medicare Access and CHIP Reauthorization Act of 2015," Pub. L. No. 114-10, 129 Stat. 87 (2015). Available at https://www.congress.gov/bill/114th-congress/house-bill/2/text.