MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3365e

Measure Title: Treatment of osteopenia or osteoporosis in men with non-metastatic prostate cancer on androgen deprivation therapy

Measure Steward: Large Urology Group Practice Association (LUGPA)

Brief Description of Measure: Men with non-metastatic prostate cancer and current or recent use of androgen deprivation therapy (ADT) and who also have a diagnosis of osteopenia or osteoporosis. The patient has been prescribed or is taking a bisphosphonate or denosumab. The patient is taking Calcium and Vitamin D supplementation, after an initial Calcium and Vitamin D level measurement. The measure scoring is proportion.

The measure focuses on this population because androgen suppression, as a treatment for prostate cancer, can cause osteoporosis. It increases bone turnover, decreases bone mineral density, and increases the risk of bone fractures in men with prostate cancer. Denosumab reduces the risk of vertebral fractures in men with prostate cancer treated with androgen deprivation therapy. Bisphosphonates increase bone mineral density, a surrogate for fracture risk, during ADT. The Endocrine Society recommends that men at high risk of fracture be treated with medication approved by regulatory agencies; at this time, alendronate, risedronate, zoledronic acid, teriparatide and denosumab for men receiving ADT for prostate cancer.

Bisphosphonates inhibit bone resorption by suppressing osteoclast activity. The addition of an osteoclast inhibitor (bisphosphonate, denosumab 60 mg every six months) in men without bone metastases who are treated with long-term ADT is indicated when the 10-year probability of hip fracture is >=3 percent or the 10-year probability of a major osteoporosis-related fracture is >=20 percent. Denosumab is a monoclonal antibody and binds to RANKL. Prevention of the RANKL/RANK interaction inhibits osteoclast formation, function, and survival, thereby decreasing bone resorption and increasing bone mass and strength in both cortical and trabecular bone. The Prolia trial studied both osteoporosis and osteopenia. At 36 months, denosumab significantly increased bone density at all measured sites (lumbar, spine, hip, femoral neck, and distal third of radius) compared with placebo. The increase in bone density was progressive over the course of time at all sites and statistically significant beginning one month after the start of treatment. Hypocalcemia must be corrected before a patient receives a bisphosphonate or denosumab. All patients should be adequately supplemented with Calcium and Vitamin D.

This measure identifies the patient with a diagnosis of osteoporosis or osteopenia who also has prostate cancer and is being placed on ADT. Osteoporosis or osteopenia treatment must start during the measurement period.

This measure is a natural progression from CMS645v1. That measure is Bone Density Evaluation for Patients with Prostate Cancer and Receiving Androgen Deprivation Therapy. If the bone density shows osteoporosis or osteopenia, and the patient is being placed on ADT, then this measure is applicable and ultimate pairing with CMS645 is desired.

Developer Rationale: Androgen suppression, as a treatment for prostate cancer, can cause osteoporosis. It increases bone turnover, decreases bone mineral density, and increases the risk of bone fractures in men with prostate cancer (Nguyen, 2015). In large population based studies, ADT was associated with a 21% to 54% relative increase in fracture risk (Shahihan, 2005), (Smith, 2006) (Smith, 2005). Denosumab reduced the risk of vertebral fractures in men with prostate cancer treated with androgen deprivation therapy (Smith, 2009). Bisphosphonates increase bone mineral density, a surrogate for fracture risk, during ADT. (Smith, 2003)(Michaelson, 2007)(Greenspan. 2007). The Endocrine Society recommends that men at high risk of fracture be treated with medication approved by regulatory agencies; at this time, alendronate, risedronate, zoledronic acid, teriparatide and denosumab for men receiving ADT for prostate cancer (Watts,2012).

Androgen Deprivation Therapy can cause osteoporosis or osteopenia in men with prostate cancer. The resulting loss of bone density can lead to skeletal related events which can decrease the quality of life for these men. Maintenance of proper bone health will improve the quality of life for men with this condition and should be recognized by those practitioners providing care for these patients.

Numerator Statement: Patients with an order for or taking bisphosphonates or denosumab and who had a Vitamin D and Calcium level completed prior to the start of treatment. Patients are also taking Calcium and Vitamin D.

Denominator Statement: The denominator equals the initial population. That is, male patients with a diagnosis of prostate cancer and osteoporosis or osteopenia. Patients with prior and/or current androgen deprivation therapy with an office visit during the measurement period. This is also the initial population.

There is no age cut off for this measure as prostate cancer can affect younger men, although it is a disease that normally occurs after the age of 40. According to the NCCN Prostate Cancer Early Detection guidelines, a cut off at 40 could miss those unfortunate patients who developed the disease in their late 20's and 30's. At the upper end, very healthy men over age 75 may choose to seek more aggressive treatment. Cancer genetics show an increased risk if the patient is a BRCA1/2 pathogenic mutation carrier which can lead to earlier detection of prostate cancers (and other cancers as well). When a family member is diagnosed with prostate cancer, another first degree relative is recommended to be screened at age 40 or 10 years prior to the age of the relative when prostate cancer was discovered, whichever is soonest.

Denominator Exclusions: Denominator Exclusions are prostate cancer with secondary metastasis to the bone and patients on comfort measures such as hospice and end of life care.

The NQF document does not include a Denominator Exception, but there is one exception which is listed in the MAT - Patient refusal of the recommendation for bisphosphonates or denosumab after the start of ADT therapy and known osteoporosis diagnosis.

Measure Type: Process
Data Source: Electronic Health Records
Level of Analysis: Clinician : Group/Practice, Clinician : Individual

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

- The developer cites the National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology Prostate Cancer version 2.2017. MS-27, MS-28. (**Level of Evidence**: Category 2A: Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate.)
    - The NCCN Guidelines Panel recommends screening and treatment for osteoporosis according to guidelines for the general population from the National Osteoporosis Foundation. The National Osteoporosis Foundation guidelines recommend:
        - 1) supplemental calcium (1200 mg daily) and vitamin D3 (800-1000 IU daily) for all men older than age 50 years; and
        - 2) additional treatment for men when the 10-year probability of hip fracture is >=3% or the 10-year probability of a major osteoporosis-related fracture is >=20%. Fracture risk can be assessed using the algorithm FRAX ®, recently released by WHO. ADT should be considered "secondary osteoporosis" using the FRAX ® algorithm. Treatment with Denosumab (60mg every 6 months), zoledronic acid (5mg IV annually, or alendronate (70mgPO weekly) is recommended when the absolute fracture risk warrants drug therapy. A baseline dual-energy x-ray absorptiometry (DEXA) scan before start of therapy and a follow-up DEXA scan after one year of therapy is recommended by the International Society of Clinical Densitometry to monitor response.

Questions for the Committee:

- *For structure, process, and intermediate outcome measures:*
  - *What is the relationship of the initiation of osteoporosis/osteopenia treatment and the bone health of patients with prostate cancer undergoing androgen deprivation therapy (ADT)?*
  - *How strong is the evidence for this relationship?*
  - *Is the evidence directly applicable to the treatment of osteopenia or osteoporosis in men with non-metastatic prostate cancer on androgen deprivation therapy*

Guidance from the Evidence Algorithm
Process measure based on systematic review (Box 3) → QQC Provided (Box 4) → Quantity: High; Quality: Moderate; Consistency: High (Box 5b) → Moderate

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer provided the following summary of data from the literature to indicate an opportunity for improvement on the treatment of osteopenia or osteoporosis in men with non-metastatic prostate cancer on androgen deprivation therapy (ADT):
  - Local studies have shown that the rates of BMD testing and/or osteoporosis treatment in ADT treatment men varied from 9-59%, with an average of less than one-quarter of ADT treated men receiving appropriate care.
- Though the data provided by the developer demonstrates an opportunity for improvement, the original study focused on the assessment and management of low bone density in inflammatory bowel disease (IBD) and performance of professional society guidelines – this is not the specific focus of the measure.
- The developer provided additional literature related to osteoporosis, assessment, and treatment.
- Testing data from the developer using two separate datasets demonstrated the following:
  - Dataset 1 (65 patients, 11 clinicians): Found a performance rate range of 0% to 87.5%, with an average performance rate of 47.91% (SD 35.27%)
  - Dataset 2: 8 patients, 1 clinician: Found a 0% performance rate.

Disparities

- The developer notes a disparity in care for this condition treatment between men and women. Per the developer, for many years there has been recognition of osteoporosis in women, especially with the hormonal loss of menopause. The ADT causes secondary osteoporosis but has often been overlooked for treatment.

Questions for the Committee:

- *Are the variations in BMD testing and osteoporosis treatments in patients (male and female) with IBD applicable to the treatment of osteopenia or osteoporosis in men with non-metastatic prostate cancer on androgen deprivation therapy (ADT)?*
- *Are you aware of evidence that disparities in the treatment of osteopenia or osteoporosis in men with non-metastatic prostate cancer on androgen deprivation therapy (ADT) exist?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence

- The evidence is level 2A suggesting that there is a high level of correlation and consensus but less direct evidence or randomized controlled trial data. The evidence suggests through large retrospective reviews that there is a relationship between ADT and the risk of developing osteoporosis/osteopenia in men diagnosed with prostate cancer. I feel that the evidence is strong and when evaluating other disease states as well as the related data in prostate cancer, generalizable across multiple clinical situations. There is also data that the use of bisphosphonates and denosumab have been demonstrated to ameliorate the negative effects of ADT on bones. NCCN does recommend screening patients with DEXA scans given the magnitude of the problem. Standard of care suggests that patients with osteoporosis/osteopenia or at high risk for bone fracture should receive Calcium and Vitamin D supplementation and, if they are at high risk from fracture, bone protectants should be added. I am not aware of additional evidence that the developers have not included in their application.
- I do not have any particular comments on this section. I do not feel qualified to judge the strength of the evidence on the relationship between initiation of osteoporosis treatment and bone health of patients. However, it does appear to be applicable to the treatment of osteoporosis in men with non-metastic prostate cancer. My initial rating for this criterion would be Moderate.
- This process measure will report the incidence of implementation of established bone health guidelines in healthy men on ADT for non-metastatic prostate cancer. There is ample evidence that ADT contributes to loss of bone density which in turn increases risk of fracture. It has further been established that a risk stratified implementation of treatment can prevent and even reverse loss of bone density.
- The NCCN Prostate Cancer Version 2.2018/March 8 2018 provides category 2A level of evidence which is "based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate". The developers note that systematic review was used (did not find that reference). I did find a literature review using PubMed only - I'm not sure if Nguyen et al. (2015) is the SR referred to or not. However, citations included RCT and recommendations from The Endocrine Society and National Osteoporosis Foundation. With the 5+ citations and the NCCN recommendation of category 2A evidence, this is moderately supported. I would like clarification as to the systematic review the developers referred to - I don't think they intended the Nguyen review to be a SR.
- The evidence underlying the NCCN guideline and citations submitted with the measure appear sufficient to support the measure and link it to preferred patient outcomes i.e. a relationship between initiation of osteoporosis/osteopenia treatment and the bone health of patients with prostate cancer undergoing androgen deprivation therapy (ADT). The evidence applies to the treatment of osteopenia or osteoporosis in men with non-metastatic prostate cancer on androgen deprivation therapy. The preliminary rating for evidence appears appropriate
- The evidence provided relates directly to the process/outcome being measured. The evidence relates to the percentage of men who receive treatment for osteoporosis or osteopenia in men with non-metastatic prostate cancer on ADT.
- Process measure with systematic review of evidence as supported by clinical practice NCCN guidelines version 3.2018, level 2A and National Osteoporosis Foundation and Endocrine Society with evidence review. Measure has moderate - high certainty the net benefit is moderate ***Please note version 3/2018 the dosing for calcium and vitamin D are slightly different than this NCCN version 22017

1b. Performance Gap

- Yes, performance data was presented. Men on ADT have between a 9-53% risk of osteoporosis. Testing and/or treatment of osteoporosis/osteopenia ranges from 9-59%. On average, less than 25% of the patients received appropriate care. These data suggest a high incidence of the problem and a high incidence of poor performance. The developers presented an analysis of two large databases (one urology group, LUGPA, and the other a radiation oncology group.) They performed chart abstraction of the electronic data. Group 1 demonstrated an average performance rate of 47.91% with a range from 0-87% among 11 clinicians. Group 2 had only 1 clinician reviewed with 0% compliance. This suggests that there is a large performance gap, especially in two practices where management of these patients is common (although the urology group is more likely to represent a practice that would administer ADT than a radiation oncology practice.) The developers also state that females are more likely to be diagnosed with bone loss and high risk of fracture due to screening

programs whereas males are less likely to be screened suggesting a gender gap in defining and treating this condition. I believe that there is a significant performance gap justifying the development of such a measure.

- I do not feel qualified to make a determination on whether variations in BMD testing & osteoporosis treatment in IBD for both men & women are applicable to treatment of men with non-metastatic prostate cancer on ADT. My initial rating would be Moderate for this criterion.
- Though perhaps not well outlined in the proposal, there is generally very low compliance or awareness of bone health measures among urologists, who predominantly administer ADT with a broad range of adherence from nothing to calcium supplements to testing and intervention. There is a need to disseminate the knowledge from other disciplines to improve quality of care for men with prostate cancer.
- The disparities seems to be primarily gender-related. Studies have primarily addressed estrogen deprivation, bone density, osteoporosis, and bisphosphonate in women and fewer studies have addressed osteoporosis in men (particularly with prostate cancer on ADT). The performance gap is based on local studies, as noted by the developers (did not find the citation for this), indicating the rates for BMD testng and/or osteoporosis treatment in men with prostate cancer with ADT ranges between 9% and 59%. Additionally, two datasets were used derived from two different facilities. One dataset includes 65 patients with 11 clinicians and the performance rate is 0% to 87.5% (over a one year period). The second dataset includes 8 patients and 1 clinician (over a one year period) with a rate of 0%. This data indicate variability in care. I am not aware of other disparities that would be pertinent to this measure.
- Their local testing showed that there is variation among physicians caring for the intended patient population. I'm unaware disparities in the treatment of osteopenia or osteoporosis in men with non-metastatic prostate cancer on androgen deprivation therapy. The preliminary rating for opportunity for improvement seems appropriate.
- Current data was provided and it demonstrates a moderate gap in treatment of men. It demonstrates a disparity in treatment of men compared to traditional treatment of women.
- Yes, with wide performance range from the 2 datasets provided. Dataset 2 very small population. There is a wide standard deviation with low performance rate indicating performance gap does exist. There is clear disparity between gender and treatment for bone health in general. Rate as moderate.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing
2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

eMeasure Technical Advisor(s) review:

Submitted measure is an	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)).

HQMF compliant eMeasure	**HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM; **OR** Submitted eMeasure contains components that cannot be represented due to limitations of HQMF or QDM and the submission explains the work around for these limitations;
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC.
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously; **OR**
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with measure developer indicates that the measure logic is feasible based on assessment by EHR vendors**;**

Complex measure evaluated by Scientific Methods Panel? ☐ **Yes** ☒ **No**
Evaluators: Staff
Evaluation of Reliability and Validity: Staff evaluation

Questions for the Committee regarding reliability and validity:

- *Do you have any concerns that the measure can be consistently implemented due to the four data elements that are not currently in structured data fields?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability (specifications)

- The data elements are clearly defined. The codes appear to be provided. All steps are logical. It is likely that the measure can be consistently implemented.
- I do not have any concerns about whether the measure can be consistently implemented.
- Four of the 16 elements do not have associated EHR fields and will require text recognition/chart review for extraction--at least on initial implementation: patient refusal, hospice status, necrosis or radiation of jaw. though all may lead to underestimation of adherence, the last two are very rare in this population.
- Two data sets (65 patients/11 clinicians and 8 patients/1clinician) were used and data abstracted from EHR (Centricity) and HQMF implemented. The developers state that all patients meeting denominator and numerator criteria were included and 100% of the data were manually analyzed. The second dataset was abstracted from an EHR that is HL7 compliant and CMS approved. All data elements were empirically tested for accuracy and correctness (100% abstraction - this was not a sample). Two extracters coordinated results (did not see inter-rater reliability).The elements from VSAC - 16 of the 20 data elements are reported in one of the EHRs. The four not currently reported can be directly linked for future reporting.
- I don't find the specifications very clear. At one point the denominator equals "male patients with a diagnosis of prostate cancer and osteoporosis or osteopenia." Subsequently "Patients with prior and/or current androgen deprivation therapy with an office visit during the measurement period. This is also the initial population." Patients are also taking Calcium and Vitamin D." Logically, it isn't evident how these are combined to define the denominator. Then the numerator is "Patients with an order for or taking bisphosphonates or denosumab and who had a Vitamin D and Calcium level completed prior to the start of treatment." Wouldn't the Vitamin D and Ca level prior to treatment be criteria for the denominator? I think the specifications for both need to be refined

to allow consistent implementation. It appears that a Denominator Exception may exist, but hasn't been fully documented.
- Data elements are clearly defined. I have no concerns that this measure can be consistently implemented.
- Feasibility may be an issue because many of the elements are not available, captured by EHR for accurate performance measure. The 2 datasets are small. 16 of 20 data elements reported.

2a2. Reliability (testing)
- The measure seems to be reliable. Obviously, all of the data elements must be captured by the clinician in order to be available and the absence of consistently captured data is a threat to reliability. But, as described by the developers, these data elements should be now or in the future, available in the EHR.
- I do not have any concerns.
- No
- Would appreciate discussion on the reliability as viewed by other providers on the committee who use EMR and particularly Centricity.
- My concern would be that different interpretations of ambiguous measure specifications might lead to different scoring. When they did their testing it seemed okay but I believe that was a single team applying the specifications as they understood them.
- I have no concerns over reliability.
- See above data elements were empirically tested for accuracy and correctness

2b1. Validity -Testing
2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)
2b4. Meaningful Differences
- There are 4 data elements of 20 proposed elements that where not readily available in an electronic form. They included: comfort measures, dental indications, oral radiation, patient refusal. The first is essential to exclude patients due to end stage disease. The next two are important to define medical exclusions for bisphosphonates. The final measure is an exclusion measure for the study. The developers did not present any evidence that there was missing data from any of the other elements. This measure would define meaningful differences between clinicians/practices. Analyses appear to produce comparable results. One feasibility concern to address is the electronic availability of DEXA scans. Usually these studies are not done in-house and the data is available to the clinician as a scanned report rather than as discreet data elements. If the diagnosis is missed as a problem, then there is not likely to be electronic evidence to confirm the diagnosis. I would ask to developers to address this concern. Overall, it does not appear that missing data will constitute a threat to validity.
- I do not have any concerns.
- As above, 4 data elements may be missing from some records, posing a small threat to validity
- Validity testing indicated 100% correlation with the value set or created with the VSAC. As noted the data elements were 3 or measurable within the EHR for 16 of the 20 elements. The developers state that all records were reviewed and all data elements were directly correlated, thus, empirically evaluated - suggesting statistical analysis of signal to noise. Missing data were noted in evident in 4 of the data sets and was then found via the narrative.
- Does this measure identifies meaningful differences about quality... yes, but it may also reflect patient preference or limitations on what is covered by their payer. Unable to ascertain whether missing data is a threat, but with regard to the possible Denominator Exception and exclusions, I don't think they've demonstrated how that might be determined.
- No concerns with validity testing. Analyses indicate that this measure identifies meaningful differences in quality based upon variations in practitioner performance.
- Validity testing showed 100% correlation to chosen value set, issues on 4 of data elements reportable in HER
-

2b2-3. Other Threats to Validity
- The exclusions are consistent with the evidence. The developers define medical reasons for ineligibility for receiving therapy with bone protectants. They also describe situations where palliative or supportive therapy would dictate that bone protectants would be unreasonable. There is no risk adjustment for this measure.
- I do not feel any patients or patient groups are inappropriately excluded from the measure. My initial rating would be Moderate for this criterion.
- Exclusions are appropriate. Conceptual relationship exists.
- The exclusions seem to be consistent with the evidence provided. Agree with preliminary analysis by staff of moderate rating for reliability and validity.

- As far as I can tell, they believe that exclusions are necessary but may not be readily implemented, which could undermine validity.
- Yes

Criterion 3. Feasibility

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

- The developer tested 20 data elements in two EHRs. Four data elements were not available in structured data fields in either EHR. On a scale from 1 to 3 where 3 is the highest score, these data elements received a score of 2 on data availability, data standards, and workflow. All other data elements are 100% available and accurate (see feasibility scorecard).
- The four data elements not available in structured data fields include:
  - Comfort Measures (denominator exclusion)
  - Dental Indications Surrounding Administration of Medications to Treat Osteoporosis in Men (numerator exclusion)
  - Oral Radiation During Medications for Osteoporosis in Men (numerator exclusion)
  - Patient Refused (denominator exception)
- The developer described a near-term path to electronic collection for the four data elements with existing SNOMED-CT and/or UMLS codes.
- The developer tested the value sets, measure logic and timing using a simulated data set (see BONNIE testing attachment).
- There are no fees or licenses associated with the use of this measure.

Questions for the Committee:

- *Are the feasibility concerns with the four data elements adequately addressed by the developer's near-term path to electronic collection?*
- *Is the data collection strategy ready to be put into operational use?*
- *Does the Committee have other feasibility concerns?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

- The data elements are routinely generated and used in the course of care delivery and are generally available as electronic elements in the EHR. There are 4 data elements of 20 proposed elements that where not readily available in an electronic form. They included: comfort measures, dental indications, oral radiation, patient refusal. The first is essential to exclude patients due to end stage disease. The next two are important to define medical exclusions for bisphosphonates. The final measure is an exclusion measure for the study. The developers propose a strategy to capture these metrics which seems feasible for sites that would like to participate in such a measure.
- I can not speak to whether the four data elements are adequately addressed by the developer's plan. I do not have any other feasibility concerns. My initial rating would be Moderate for this criterion.
- As above, four exclusion elements are not routinely generated and will require text recognition or chart review.
- I don't have concerns at this time about feasibility. Tested 20 data elements in two EHRs (one was Centricity). of the 20, 4 were not available in the data fields for either EHR. Developers discussed how to address the 4 data elements not included in the current fields.
- The feasibility concerns with the four data elements are addressed by the developer's near-term path to electronic collection, but I'm skeptical that the suggestion for a template would find widespread implementation for the purpose of using this measure. Hence, I wouldn't say that the data collection strategy is really ready to be put into operational use.

- Comfort Measures (denominator exclusion); Dental Indications Surrounding Administration of Medications to Treat Osteoporosis in Men (numerator exclusion); Oral Radiation During Medications for Osteoporosis in Men (numerator exclusion); Patient Refused (denominator exception)
- Dataset 2 feasibility 77% only 11 data elements allowed limiting analysis per the submitter I am unaware if EHR can cover all the numerator exclusions i.e. inflammation of gums, prior osteonecrosis of jaw, etc.

Criterion 4: Usability and Use

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No** ☐ **UNCLEAR**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- The measure is currently not in use in any accountability or payment programs.
- The measure is planned for use in MIPS.
- The measure is also planned for use in regulatory and accreditation programs, internal quality improvement, and external quality improvement benchmarking.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- This measure was presented in the draft form to the annual meeting of the Large Urology Group Practice Association (LUGPA). It was discussed and specific elements reviewed. Feedback was supportive of the inclusion of urology measures in MIPS.
- This measure was submitted for 2017 Measures Under Consideration and feedback was obtained from CMS for further development. CMS suggested further development and BONNIE testing, which was done. The input was incorporated into the measure and further testing was completed.
- CMS also requested that the developer consider pairing this measure with measure CMS645v1, Bone Density Evaluation for Patients with Prostate Cancer and Receiving Androgen Deprivation Therapy, which is also stewarded by Oregon Urology and is not NQF endorsed.

***Questions for the Committee**:*

- *How has the measure been vetted in real-world settings by those being measured or others?*

Preliminary rating for Use: ☐ **Pass** ☒ **No Pass**

RATIONALE: This is a new measure and is not required to be in an accountability program or publicly reported at the time of initial endorsement.

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. Progress on improvement is not required for initial endorsement unless available.

- The developer states that implementing and reporting this measure will assist clinicians to look at their performance from month to month and determine what needs to be done to improve performance. The developer describes the quality improvement processes that should reduce risk and maintain health in men with osteoporosis.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- The developer reported no unexpected findings.

Potential harms

- The developer did not report any potential harms

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for Usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Use

- The developers described multiple different responses to the measure and report that reviewers uniformly approved the measure and the data presented.
- I can not say how or if the measure has been vetted in a real-world setting.
- New measure--Not currently in use or planned for public reporting but has been proposed to CMS as a urologic MIPS measure.
- The measure provides greater benefit than harm. The measure can be reported monthly for clinicians to view and potentially improve their own performance.
- The implementation plan seems credible, but previously noted difficulties with specifications/reliability/validity could pose significant obstacles I'm afraid.
- The measure is not being publicly reported but is not required to be. The measure was presented to the LUGPA, and submitted for feedback from CMS
- New measure, n/a

4b. Usability

- There do not appear to be any actual or unintended consequences from using this measure. The benefits of administering appropriate therapy to men diagnosed with prostate cancer on ADT appear to outweigh any risks.
- I would think the results could be used to further the goal of high-quality healthcare by providing evidence for additional treatment for osteoporosis when ADT is utilized. I would think the benefits would outweigh the unintended consequences
- The proposed measure has high potential to increase awareness of and adherence to well established guidelines for bone health management in men on ADT. Potential harm will be the increased cost burden of testing and treatment, will be balanced by decreases costs related to fracture with subsequent health care needs.
- See above comment
- The measure hasn't been vetted in real-world settings. If properly implement it does seem on track to provide useful data on quality care for this patient population. Unintended consequences seem unlikely.
- Yes, the performance results can improve the goal of high quality. No unintended consequences or harm.
- New measure n/a; plan future use in the real world setting, should vet this with other groups as priority. understand this is being paired with another measure but not sure about implementation and priority to other groups, would need buy in from other stakeholders

Criterion 5: Related and Competing Measures

Related or competing measures

- There are no related or competing measures associated with this measure.

Committee pre-evaluation comments

Criterion 5: Related and Competing Measures

N/A

Public and member comments

Comments and Member Support/Non-Support Submitted as of: Month/Day/Year

- **Of the XXX NQF members who have submitted a support/non-support choice:**
  - XX support the measure
  - YY do not support the measure

Measure Number: 3365e

Measure Title: Treatment of osteopenia or osteoporosis in men with non-metastatic prostate cancer on androgen deprivation therapy

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

Instructions for filling out this form:

- Please complete this form for each measure you are evaluating.
- Please pay close attention to the skip logic directions. ***Directives that require you to skip questions are marked in red font.***
- If you are unable to check a box, please highlight or shade the box for your response.
- You must answer the "overall rating" item for both Reliability and Validity. Also, be sure to answer the composite measure question at the end of the form if your measure is a composite.
- For several questions, we have noted which sections of the submission documents you should ***REFERENCE*** and provided ***TIPS*** to help you answer them.
- ***It is critical that you explain your thinking/rationale if you check boxes that require an explanation.*** Please add your explanation directly below the checkbox in a different font color***.*** Also, feel free to add additional explanation, even if you select a checkbox where an explanation is not requested (if you do so, please type this text directly below the appropriate checkbox).
- ***Please refer to the Measure Evaluation Criteria and Guidance document (pages 18-24) and the 2-page Key Points document when evaluating your measures***. This evaluation form is an adaptation of Alogorithms 2 and 3, which provide guidance on rating the Reliability and Validity subcriteria.
- ***Remember*** that testing at either the data element level **OR** the measure score level is accepted for some types of measures, but not all (e.g., instrument-based measures, composite measures), and therefore, the embedded rating instructions may not be appropriate for all measures.
- ***Please base your evaluations solely on the submission materials provided by developers.*** NQF strongly discourages the use of outside articles or other resources, even if they are cited in the submission materials. If you require further information or clarification to conduct your evaluation, please communicate with NQF staff (methodspanel@qualityforum.org).

RELIABILITY

1. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?
   REFERENCE: "MIF_xxxx" document
   NOTE*: NQF staff will conduct a separate, more technical, check of eMeasure (eCQM) specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*
   TIPS*: Consider the following: Are all the data elements clearly defined? Are all appropriate codes included? Is the logic or calculation algorithm clear? Is it likely this measure can be consistently implemented?*
   ☒Yes (go to Question #2)
   ☐No (please explain below, and go to Question #2) NOTE that even though ***non-precise specifications should result in an overall LOW rating for reliability***, we still want you to look at the testing results.

2. Was empirical reliability testing (at the data element or measure score level) conducted using statistical tests with the measure as specified?

   REFERENCE: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2
   ***TIPS**: Check the "NO" box below if: only descriptive statistics are provided; only describes process for data management/cleaning/computer programming; testing does not match measure specifications (i.e., data source, level of analysis, included patients, etc.)*

   ☐Yes (go to Question #3)

   ☒No, there is reliability testing information, but *not* using statistical tests and/or not for the measure as specified **OR** there is no reliability testing (please explain below, skip Questions #3-8, then go to Question #9)

   Empirical validity testing of patient-level data conducted. Per NQF guidance, use rating from patient-level data element validity testing.

3. Was reliability testing conducted with computed performance measure scores for each measured entity?

   REFERENCE: "Testing attachment_xxx", section 2a2.1 and 2a2.2
   ***TIPS**: Answer no if: only one overall score for all patients in sample used for testing patient-level data*

   ☐Yes (go to Question #4)

   ☐No (skip Questions #4-5 and go to Question #6)

4. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? *NOTE: If multiple methods used, at least one must be appropriate.*

   REFERENCE: Testing attachment, section 2a2.2
   ***TIPS**: Examples of appropriate methods include signal-to-noise analysis (e.g. Adams/RAND tutorial); random split-half correlation; other accepted method with description of how it assesses reliability of the performance score.*

   ☐Yes (go to Question #5)

   ☐No (please explain below, then go to question #5 and rate as INSUFFICIENT)

5. **RATING (score level)** - What is the level of certainty or confidence that the performance measure scores are reliable?

   REFERENCE: Testing attachment, section 2a2.2
   ***TIPS**: Consider the following: Is the test sample adequate to generalize for widespread implementation? Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

   ☐High (go to Question #6)

   ☐Moderate (go to Question #6)

   ☐Low (please explain below then go to Question #6)

   ☐Insufficient (go to Question #6)

6. Was reliability testing conducted with patient-level data elements that are used to construct the performance measure?

   REFERENCE: Testing attachment, section 2a2.
   ***TIPS**: Prior reliability studies of the same data elements may be submitted; if comparing abstraction to "authoritative source/gold standard" go to Question #9)*

   ☐Yes (go to Question #7)

   ☐No (if there is score-level testing that you rated something other than INSUFFICIENT in Question #5, skip questions #7-9, then go to Question #10 (OVERALL RELIABILITY); otherwise, skip questions #7-8 and go to Question #9)

7. Was the method described and appropriate for assessing the reliability of ALL critical data elements?
   REFERENCE: Testing attachment, section 2a2.2
   TIPS*: For example: inter-abstractor agreement (ICC, Kappa); other accepted method with description of how it assesses reliability of the data elements*
   Answer no if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)
   ☐Yes (go to Question #8)
   ☐No (if no, please explain below, then go to Question #8 and rate as INSUFFICIENT)

8. **RATING (data element)** – Based on the reliability statistic and scope of testing (number and representativeness of patients and entities), what is the level of certainty or confidence that the data used in the measure are reliable?
   REFERENCE: Testing attachment, section 2a2
   TIPS*: Consider the following: Is the test sample adequate to generalize for widespread implementation? Can data elements be collected consistently?*
   ☐Moderate (skip Question #9 and go to Question #10 (OVERALL RELIABILITY); if score-level testing was NOT conducted, rate Question #10 as MODERATE)
   ☐Low (skip Question #9 and go to Question #10 (OVERALL RELIABILITY); if score-level testing was NOT conducted, rate Question #10 (OVERALL RELIABILITY) as LOW)
   ☐Insufficient (go to Question #9)

9. Was **empirical VALIDITY testing** of patient-level data conducted?
   REFERENCE: testing attachment section 2b1.
   NOTE: Skip this question if empirical reliability testing was conducted and you have rated Question #5 and/or #8 as anything other than INSUFFICIENT)
   TIP: *You should answer this question ONLY if score-level or data element reliability testing was NOT conducted or if the methods used were NOT appropriate. For most measures, NQF will accept data element validity testing in lieu of reliability testing—but* ***check with NQF staff before proceeding, to verify*****.**
   ☒Yes (go to Question #10 and answer using your rating from data element validity testing – Question #23)
   ☐No (please explain below, go to Question #10 (OVERALL RELIABILITY) and rate it as INSUFFICIENT. Then go to Question #11.)

OVERALL RELIABILITY RATING

10. **OVERALL RATING OF RELIABILITY** taking into account precision of specifications (see Question #1) and all testing results:
    ☐High (NOTE: Can be HIGH only if score-level testing has been conducted)
    ☒Moderate (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)
    ☐Low (please explain below) [NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete]
    ☐Insufficient (please explain below) [NOTE: For most measure types, testing at both the score level and the data element level is not required, but check with NQF staff]

VALIDITY

Assessment of Threats to Validity

11. Were potential threats to validity that are relevant to the measure empirically assessed ()?

REFERENCE: Testing attachment, section 2b2-2b6

***TIPS**: Threats to validity that should be assessed include: exclusions; need for risk adjustment; ability to identify statistically significant and meaningful differences; multiple sets of specifications; missing data/nonresponse.*

☒Yes (go to Question #12)

☐No (please explain below and then go to Question #12) [NOTE that ***non-assessment of applicable threats should result in an overall INSUFFICENT rating for validity***]

12. Analysis of potential threats to validity: Any concerns with measure exclusions?

REFERENCE: Testing attachment, section 2b2.

***TIPS**: Consider the following: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)? If patient preference (e.g., informed decisionmaking) is a basis for exclusion, does it impact performance and if yes, is the measure specified so that the information about patient preference and the effect on the measure is transparent?*

☒Yes (please explain below then go to Question #13)

☐No (go to Question #13)

☐Not applicable (i.e., there are no exclusions specified for the measure; go to Question #13)

Exclusions based on clinical evidence. However, Comfort Measures (denominator exclusion), Patient Refusal (denominator exception), and Dental Indications Surrounding Administration of Medications to Treat Osteoporosis in Men and Oral radiation During Medications for Osteoporosis in Men (two numerator exclusions) were not available in structured data fields in the two EHRs tested. The measure developer states they will track excluded patients once they fully implement the measure.

13. Analysis of potential threats to validity: Risk-adjustment (this applies to <u>all</u> outcome, cost, and resource use measures and "NOT APPLICABLE" is not an option for those measures; the risk-adjustment questions (13a-13c, below) also may apply to other types of measures)

REFERENCE: Testing attachment, section 2b3.

13a. Is a conceptual rationale for social risk factors included? ☐Yes ☐No

13b. Are social risk factors included in risk model? ☐Yes ☐No

13c. Any concerns regarding the risk-adjustment approach?

***TIPS**: Consider the following: **If measure is risk adjusted**: If the developer asserts there is **no conceptual basis** for adjusting this measure for social risk factors, do you agree with the rationale? Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented? Are all of the risk adjustment variables present at the start of care (if not, do you agree with the rationale)? If social risk factors are not included in the risk-adjustment approach, do you agree with the developer's decision? Is an appropriate risk-adjustment strategy included in the measure (e.g., adequate model discrimination and calibration)? Are all statistical model specifications included, including a "clinical model only" if social risk factors are included in the final model? If a measure is NOT risk-adjusted, is a justification for **not risk adjusting** provided (conceptual and/or empirical)? Is there any evidence that contradicts the developer's rationale and analysis for not risk-adjusting?*

☐Yes (please explain below then go to Question #14)

☐No (go to Question #14)

☒Not applicable (e.g., this is a structure or process measure that is not risk-adjusted; go to Question #14)

14. Analysis of potential threats to validity: Any concerns regarding ability to identify meaningful differences in performance or overall poor performance?

REFERENCE: Testing attachment, section 2b4.

☐Yes (please explain below then go to Question #15)

☒No (go to Question #15)
This is a new measure with data from pilot testing.

15. Analysis of potential threats to validity: Any concerns regarding comparability of results if multiple data sources or methods are specified?
REFERENCE: Testing attachment, section 2b5.
☐Yes (please explain below then go to Question #16)
☐No (go to Question #16)
☒Not applicable (go to Question #16)

16. Analysis of potential threats to validity: Any concerns regarding missing data?
REFERENCE: Testing attachment, section 2b6.
☒Yes (please explain below then go to Question #17)
☐No (go to Question #17)
Four data elements, Comfort Measures (denominator exclusion), Patient Refusal (denominator exception), and Dental Indications Surrounding Administration of Medications to Treat Osteoporosis in Men and Oral Radiation During Medications for Osteoporosis in Men (two numerator exclusions) were not available in structured data fields in the two EHRs tested. Measure developer found missing data elements in the narrative format. If data elements not found in narrative format, then data considered missing – no empirical analysis of missing data provided.

Assessment of Measure Testing

17. Was empirical validity testing conducted using the measure as specified and with appropriate statistical tests?
REFERENCE: Testing attachment, section 2b1.
***TIPS**: Answer no if: only face validity testing was performed; only refer to clinical evidence; only descriptive statistics; only describe process for data management/cleaning/computer programming; testing does not match measure specifications (i.e. data, eMeasure, level, setting, patients).*
☒Yes (go to Question #18)
☐No (please explain below, then skip Questions #18-23 and go to Question #24)

18. Was validity testing conducted with computed performance measure scores for each measured entity?
REFERENCE: Testing attachment, section 2b1.
***TIPS**: Answer no if: one overall score for all patients in sample used for testing patient-level data.*
☐Yes (go to Question #19)
☒No (please explain below, then skip questions #19-20 and go to Question #21)

19. Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?
REFERENCE: Testing attachment, section 2b1.
***TIPS**: For example: correlation of the performance measure score on this measure and other performance measures; differences in performance scores between groups known to differ on quality; other accepted method with description of how it assesses validity of the performance score*
☐Yes (go to Question #20)
☐No (please explain below, then go to Question #20 and rate as INSUFFICIENT)

20. **RATING (measure score)** - Based on the measure score results (significance, strength) and scope of testing (number of measured entities and representativeness) and analysis of potential threats, what is the level of certainty or confidence that the performance measure scores are a valid indicator of quality?

☐High (go to Question #21)
☐Moderate (go to Question #21)
☐Low (please explain below then go to Question #21)
☐Insufficient (go to Question #21)

21. Was validity testing conducted with patient-level data elements?
REFERENCE: Testing attachment, section 2b1.
***TIPS**: Prior validity studies of the same data elements may be submitted*
☒Yes (go to Question #22)
☐No (if there is score-level testing that you rated something other than INSUFFICIENT in Question #20, skip questions #22-25, and go to Question #26 (OVERALL VALIDITY); otherwise, skip questions #22-23 and go to Question #24)

22. Was the method described and appropriate for assessing the accuracy of ALL critical data elements? *NOTE that data element validation from the literature is acceptable.*
***REFERENCE:** Testing attachment, section 2b1.*
***TIPS**: For example: Data validity/accuracy as compared to authoritative source- sensitivity, specificity, PPV, NPV; other accepted method with description of how it assesses validity of the data elements.*
Answer No if: only assessed percent agreement; did not assess separately for all data elements (at least numerator, denominator, exclusions)
☒Yes (go to Question #23)
☐No (please explain below, then go to Question #23 and rate as INSUFFICIENT)

23. **RATING (data element)** - Based on the data element testing results (significance, strength) and scope of testing (number and representativeness of patients and entities) and analysis of potential threats, what is the level of certainty or confidence that the data used in the measure are valid?
☒Moderate (skip Questions #24-25 and go to Question #26)
☐Low (please explain below, skip Questions #24-25 and go to Question #26)
☐Insufficient (go to Question #24 only if no other empirical validation was conducted OR if the measure has not been previously endorsed; otherwise, skip Questions #24-25 and go to Question #26)

The measure developer tested 20 data elements for 73 patients and 12 providers in 2 EHRs. Four data elements (see comments in exclusions and missing data above) were not available in structured data fields in either EHR. The measure developer stated the four data elements could be operational within 6 months. The data elements needed to calculate the numerator, denominator, and additional exclusions are 100% accurate and available – see feasibility scorecard.

24. Was face validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from the measure as specified can be used to distinguish good and poor quality?
NOTE: If appropriate empirical testing has been conducted, then evaluation of face validity is not necessary; you should skip this question and Question 25, and answer Question #26 based on your answers to Questions #20 and/or #23]
REFERENCE: Testing attachment, section 2b1.
***TIPS**: Answer no if: focused on data element accuracy/availability/feasibility/other topics; the degree of consensus and any areas of disagreement not provided/discussed.*
☐Yes (go to Question #25)
☐No (please explain below, skip question #25, go to Question #26 (OVERALL VALIDITY) and rate as INSUFFICIENT)

25. **RATING (face validity)** - Do the face validity testing results indicate substantial agreement that the performance measure score from the measure as specified can be used to distinguish quality AND potential threats to validity are not a problem, OR are adequately addressed so results are not biased?

REFERENCE: Testing attachment, section 2b1.

***TIPS**: Face validity is no longer accepted for maintenance measures unless there is justification for why empirical validation is not possible and you agree with that justification.*

☐Yes (if a NEW measure, go to Question #26 (OVERALL VALIDITY) and rate as MODERATE)

☐ Yes (if a MAINTENANCE measure, do you agree with the justification for not conducting empirical testing? If no, go to Question #26 (OVERALL VALIDITY) and rate as INSUFFICIENT; otherwise, rate Question #26 as MODERATE)

☐No (please explain below, go to Question #26 (OVERALL VALIDITY) and rate AS LOW)

OVERALL VALIDITY RATING

26. **OVERALL RATING OF VALIDITY** taking into account the results and scope of all testing and analysis of potential threats.

☐High (NOTE: Can be HIGH only if score-level testing has been conducted)

☒Moderate (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☐Low (please explain below) [NOTE: Should rate LOW if you believe that there are threats to validity and/or threats to validity were not assessed]

☐Insufficient (if insufficient, please explain below) [NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT—please check with NQF staff if you have questions.]

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Treatment of osteopenia or osteoporosis in men with non-metastatic prostate cancer on androgen deprivation therapy.
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: 4/12/2018

Instructions

- *Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in

such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome

☐ Outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Male Patient with Prostate Cancer and Osteoporosis or Osteopenia on Androgen Deprivation Therapy (ADT) Who Are Receiving Treatment for Bone Health.

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

The patient has had a DXA scan (bone density scan) and is determined to have osteoporosis or osteopenia. The DXA scan was done because the patient has non-metastatic prostate cancer and is on androgen deprivation therapy (ADT). The patient must have a calcium and vitamin D blood level determined and corrected if abnormal. They then are prescribed a bisphosphonate or denosumab for treatment of osteoporosis. Drug prescribed and possible exclusions for taking the drugs are poor dentition, osteonecrosis of the jaw present or by history, planned radiation to the jaw or history of jaw radiation and sensitivity to the drug. For bisphosphonate patients, exclusions are hypocalcemia not corrected, creatinine clearance below 35mL/min, history of Barrett's esophagus and known drug hypersensitivity. Patients to be excluded from the measure would be those with secondary metastasis to the bone and those on comfort measures such as end of life care and hospice care.

The outcome being measured is the appropriate initiation of osteoporosis/osteopenia treatment to this patient taking a high risk drug.

1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

Several patient reports are included with this measure. First is the DXA scan which is meaningful for men with prostate cancer and on ADT. The risk for bone demineralization is large. A physician/provider discusses the results of the DXA. Any patient on ADT should have a Vitamin D and Calcium level checked and be placed on Calcium and Vitamin D as needed and at appropriate doses to supplement dietary intake. The risk of demineralization and thereby risk of a skeletal event are greater. The patient with osteoporosis or osteopenia is then placed on a bisphosphonate or denosumab to prevent or reduce demineralization. The provider/physician explains the entire cycle of bone demineralization and the risks involved with the medications. A number of men have not previously been concerned about osteoporosis or bone demineralization. This measure is very important for men on ADT as the androgen deprivation therapy can be the cause for loss of bone density. Historically, bone density has not been discussed to a great degree with men, unless they have an osteoporosis related fracture which unfortunately can markedly change their quality of life.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

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

x Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**National comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology Prostate Cancer version 2.2017** **February 21, 2017.** **MS-27, MS-28** https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	The NCCN Guidelines Panel recommends screening and treatment for osteoporosis according to guidelines for the general population from the National Osteoporosis Foundation. The National Osteoporosis Foundation guidelines include: 1) supplemental calcium (1200 mg daily) and vitamin D3 (800-1000 IU daily) for all men older than age 50 years; and 2) additional treatment for men when the 10-year probability of hip fracture is >=3% or the 10-year probability of a major osteoporosis-related fracture is >=20%. Fracture risk can be assessed using the algorithm FRAX ®, recently released by WHO. ADT should be considered "secondary osteoporosis" using the FRAX ® algorithm. Currently, treatment with deonsumab (60mg every 5 months), zoledronic acid (5mg IV annually, or alendronate (70mgPO weekly) is recommended when the absolute fracture risk warrants drug therapy. A baseline dual-energy x-ray absorptiometry (DEXA) scan before start of therapy and a follow-up DEXA scan after one year of therapy is recommended by the International Society of Clinical Densitometry to monitor response.

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	NCCN Category of Evidence and Consensus Category 2A: Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate,
Provide all other grades and definitions from the evidence grading system	NCCN Categories of Evidence and Consensus Category 1: Based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate. Category 2B: Based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate. Category 3: Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate. All recommendations are category 2A unless otherwise noted.
Grade assigned to the **recommendation** with definition of the grade	NCCN Category of Evidence and Consensus Category 2A: Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate,
Provide all other grades and definitions from the recommendation grading system	NCCN Categories of Evidence and Consensus Category 1: Based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate. Category 2B: Based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate. Category 3: Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate. All recommendations are category 2A unless otherwise noted.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Studies on bisphosphonates and denosumab (Prolia). There are greater than 5 studies on bisphosphonates and denosumab. Many initial studies were on women. The denosumab studies focused on the Prolia dosage and frequency of denosumab for men and women. The Xgeva study focused on the higher dosage and frequency of denosumab for men. All of the studies (as well as this measure) discuss and outline the potential harmful outcomes or risks and evaluation of risks to be considered prior to administration of these medications. Table 2 – Evaluation of Quantity, Quality, and Consistency of Body of Evidence for Structure, Process, and Intermediate Outcome Measures (from the Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement, August 2017. Evaluation of this measure within this Table place the definition as High and the Quantity of Body of Evidence as 5+ studies. The Quality of Body of Evidence is in agreement with this measure, that is randomized controlled trials providing direct evidence for the specific measure focus, with adequate size to obtain precise estimates of effect, and without serious flaws that introduce bias. Lastly, the Consistency of Results of Body of Evidence relates to this measure as the estimates of clinically/practically meaningful benefits and harms to patients are consistent in direction and similar magnitude across the preponderance of studies in the body of evidence.
Estimates of benefit and consistency across studies	Bisphosphonates increase bone mineral density, a surrogate for fracture risk, during ADT. Denosumab binds to and inhibits the receptor activator of RANKL to blunt osteoclast function and delays generalized bone resorption and local bone destruction.

What harms were identified?	Osteonecrosis of the jaw, atrial fibrillation, esophageal irritation, hypocalcemia, risk with poor kidney function
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	No new studies since the inception of the measure – January 2017.

1a.4 OTHER SOURCE OF EVIDENCE

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

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

The international Osteoporosis Foundation and the Endocrine Society clinical Practice Guidelines both support the same guidelines which are clearly outlined by NCCN. Bone loss with ADT is very real and can be prevented or reversed with Calcium, vitamin D and either a bisphosphonate or denosumab. Osteoporosis or osteopenia must be evaluated and treated in men with this condition.

1a.4.2 What process was used to identify the evidence?

Literature review and use of clinically accepted urology guidelines.

1a.4.3. Provide the citation(s) for the evidence.

https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf
https://www.ncbi.nlm.nih.gov/pubmed/12771706
https://academic.oup.com/jcem/article-lookup/doi/10.1210/jc.2011-3045
https://www.ncbi.nlm.nih.gov/pubmed/25097095
https://www.ncbi.nlm.nih.gov/pubmed/17369566

Measure Information

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

Brief Measure Information

NQF #: 3365e
Corresponding Measures:
De.2. **Measure Title:** Treatment of osteopenia or osteoporosis in men with non-metastatic prostate cancer on androgen deprivation therapy
Co.1.1. **Measure Steward:** Large Urology Group Practice Association
De.3. **Brief Description of Measure:** Men with non-metastatic prostate cancer and current or recent use of androgen deprivation therapy (ADT) and who also have a diagnosis of osteopenia or osteoporosis. The patient has been prescribed or is taking a bisphosphonate or denosumab. The patient is taking Calcium and Vitamin D supplementation, after an initial Calcium and Vitamin D level measurement. The measure scoring is proportion.
The measure focuses on this population because androgen suppression, as a treatment for prostate cancer, can cause osteoporosis. It increases bone turnover, decreases bone mineral density, and increases the risk of bone fractures in men with prostate cancer. Denosumab reduces the risk of vertebral fractures in men with prostate cancer treated with androgen deprivation therapy. Bisphosphonates increase bone mineral density, a surrogate for fracture risk, during ADT. The Endocrine Society recommends that men at high risk of fracture be treated with medication approved by regulatory agencies; at this time, alendronate, risedronate, zoledronic acid, teriparatide and denosumab for men receiving ADT for prostate cancer.
Bisphosphonates inhibit bone resorption by suppressing osteoclast activity. The addition of an osteoclast inhibitor (bisphosphonate, denosumab 60 mg every six months) in men without bone metastases who are treated with long-term ADT is indicated when the 10-year probability of hip fracture is >=3 percent or the 10-year probability of a major osteoporosis-related fracture is >=20 percent. Denosumab is a monoclonal antibody and binds to RANKL. Prevention of the RANKL/RANK interaction inhibits osteoclast formation, function, and survival, thereby decreasing bone resorption and increasing bone mass and strength in both cortical and trabecular bone. The Prolia trial studied both osteoporosis and osteopenia. At 36 months, denosumab significantly increased bone density at all measured sites (lumbar, spine, hip, femoral neck, and distal third of radius) compared with placebo. The increase in bone density was progressive over the course of time at all sites and statistically significant beginning one month after the start of treatment. Hypocalcemia must be corrected before a patient receives a bisphosphonate or denosumab. All patients should be adequately supplemented with Calcium and Vitamin D.
This measure identifies the patient with a diagnosis of osteoporosis or osteopenia who also has prostate cancer and is being placed on ADT. Osteoporosis or osteopenia treatment must start during the measurement period.
This measure is a natural progression from CMS645v1. That measure is Bone Density Evaluation for Patients with Prostate Cancer and Receiving Androgen Deprivation Therapy. If the bone density shows osteoporosis or osteopenia, and the patient is being placed on ADT, then this measure is applicable and ultimate pairing with CMS645 is desired.
1b.1. **Developer Rationale:** Androgen suppression, as a treatment for prostate cancer, can cause osteoporosis. It increases bone turnover, decreases bone mineral density, and increases the risk of bone fractures in men with prostate cancer (Nguyen, 2015). In large population based studies, ADT was associated with a 21% to 54% relative increase in fracture risk (Shahihan, 2005), (Smith, 2006) (Smith, 2005). Denosumab reduced the risk of vertebral fractures in men with prostate cancer treated with androgen deprivation therapy (Smith, 2009). Bisphosphonates increase bone mineral density, a surrogate for fracture risk, during ADT. (Smith, 2003)(Michaelson, 2007)(Greenspan. 2007). The Endocrine Society recommends that men at high risk of fracture be

treated with medication approved by regulatory agencies; at this time, alendronate, risedronate, zoledronic acid, teriparatide and denosumab for men receiving ADT for prostate cancer (Watts,2012).
Androgen Deprivation Therapy can cause osteoporosis or osteopenia in men with prostate cancer. The resulting loss of bone density can lead to skeletal related events which can decrease the quality of life for these men. Maintenance of proper bone health will improve the quality of life for men with this condition and should be recognized by those practitioners providing care for these patients.

S.4. **Numerator Statement:** Patients with an order for or taking bisphosphonates or denosumab and who had a Vitamin D and Calcium level completed prior to the start of treatment. Patients are also taking Calcium and Vitamin D.
S.6. **Denominator Statement:** The denominator equals the initial population. That is, male patients with a diagnosis of prostate cancer and osteoporosis or osteopenia. Patients with prior and/or current androgen deprivation therapy with an office visit during the measurement period. This is also the initial population.
There is no age cut off for this measure as prostate cancer can affect younger men, although it is a disease that normally occurs after the age of 40. According to the NCCN Prostate Cancer Early Detection guidelines, a cut off at 40 could miss those unfortunate patients who developed the disease in their late 20's and 30's. At the upper end, very healthy men over age 75 may choose to seek more aggressive treatment. Cancer genetics show an increased risk if the patient is a BRCA1/2 pathogenic mutation carrier which can lead to earlier detection of prostate cancers (and other cancers as well). When a family member is diagnosed with prostate cancer, another first degree relative is recommended to be screened at age 40 or 10 years prior to the age of the relative when prostate cancer was discovered, whichever is soonest.
S.8. **Denominator Exclusions:** Denominator Exclusions are prostate cancer with secondary metastasis to the bone and patients on comfort measures such as hospice and end of life care.
The NQF document does not include a Denominator Exception, but there is one exception which is listed in the MAT - Patient refusal of the recommendation for bisphosphonates or denosumab after the start of ADT therapy and known osteoporosis diagnosis.

De.1. **Measure Type:** Process
S.17. **Data Source:** Electronic Health Records
S.20. **Level of Analysis:** Clinician : Group/Practice, Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

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

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

De.4. **IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?** Ultimately, it is desired for this measure to be paired with CMS645. CMS645 is active within MIPS2018 but has not yet been submitted for NQF endorsement.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

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

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
NQF_evidence_attachment_Sep2017_Osteoporosis_measure_04122018.docx
1a.1 <u>For Maintenance of Endorsement:</u> Is there new evidence about the measure since the last update/submission?
Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*
If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.
Androgen suppression, as a treatment for prostate cancer, can cause osteoporosis. It increases bone turnover, decreases bone mineral density, and increases the risk of bone fractures in men with prostate cancer (Nguyen, 2015). In large population based studies, ADT was associated with a 21% to 54% relative increase in fracture risk (Shahihan, 2005), (Smith, 2006) (Smith, 2005). Denosumab reduced the risk of vertebral fractures in men with prostate cancer treated with androgen deprivation therapy (Smith, 2009). Bisphosphonates increase bone mineral density, a surrogate for fracture risk, during ADT. (Smith, 2003)(Michaelson, 2007)(Greenspan. 2007). The Endocrine Society recommends that men at high risk of fracture be treated with medication approved by regulatory agencies; at this time, alendronate, risedronate, zoledronic acid, teriparatide and denosumab for men receiving ADT for prostate cancer (Watts,2012).
Androgen Deprivation Therapy can cause osteoporosis or osteopenia in men with prostate cancer. The resulting loss of bone density can lead to skeletal related events which can decrease the quality of life for these men. Maintenance of proper bone health will improve the quality of life for men with this condition and should be recognized by those practitioners providing care for these patients.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis**. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*
N/A

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
Male bone health has not been addressed in the quality measures prior to the MIPS 2018 inclusion of Bone density evaluation for patients with prostate cancer and receiving androgen deprivation therapy- CMS645v1. This measure takes the next step in treating known osteoporosis/osteopenia in male patients who are undergoing a necessary treatment for prostate cancer that is known to result in bone loss. Initial data for this measure shows considerable variation or less than optimal performance among providers. Over several years, this has been an emphasis for physicians caring for these patients but is not regularly addressed in a formal fashion by all physicians. There is marked room for improvement which will result in better care, reduced fractures and bone density loss and the resulting pain from those conditions.

According to the International Osteoporosis Foundation in their report of 2016, there are gaps and solutions in bone health. Despite the global threat posed by fragility fractures, and the availability of safe and cost-effective therapies that could reduce the number of fractures, GAPS in CARE are preventing millions of at-risk individuals from being diagnosed and treated worldwide.

(Within CARE GAP 1 (Secondary fracture prevention): Highly effective osteoporosis treatments substantially reduce fracture risk but are often not routinely offered to fragility fracture sufferers.

Within CARE GAP 2 (Osteoporosis induced by medicines): Many widely used medicines have been associated with decreases in bone mineral density (BMD) and/or increased fracture incidence. Three of the more commonly used agents which significantly affect bone health are glucocorticoids for the treatment of a range of conditions, androgen deprivation therapy (ADT) for treatment of prostate cancer in men, and aromatase inhibitors for treatment of hormone receptor-positive breast cancer in women.

CARE Gap 3 (Diseases associated with osteoporosis): Approximately half of men diagnosed with prostate cancer will receive ADT after diagnosis. A meta-analysis reported that between 9 % and 53% of survivors had osteoporosis. Local studies have shown

that the rates of BMD testing and/or osteoporosis treatment in ADT treatment men varied from 9-59%, with an average of less than one-quarter of ADT treated men receiving appropriate care.

Within CARE GAP 4 (Primary fracture prevention for individuals at high risk of fracture): Several health systems have implemented systematic approaches to primary fracture prevention targeted at high risk individuals in parallel to secondary prevention. More healthy systems need to follow these examples.

CARE GAP 5 (Suboptimal communication and low public awareness – The importance of staying on prescribed treatment): It has been estimated that improved adherence in the USA would reduce fracture rates by 25%, equating to ca. 300,000 fewer fractures per year and generating savings of US$3 billion. Within CARE GAP 8 (Impeding access and reimbursement to osteoporosis assessment and treatment): Only partial reimbursement, or restrictive criteria for reimbursement, of diagnostic testing and drug treatment. In the USA, for example, a major drop in reimbursement for DXA testing in the office setting has led to a drop in the number of providers and more than 1 million fewer DXAs performed. In some countries or regions within a country, not all osteoporosis drugs are reimbursed, effectively limiting treatment options for individuals in need.1

In the September 2014 issue of the Journal of Bone and Mineral Research, Solomon and colleagues report on the uptake of osteoporosis medications in the year following hip fracture in a large retrospective analysis of nearly 100,000 men and women aged 50 years or more who were hospitalized for hip fracture over a period of 1 year. The estimated probability of receiving osteoporosis medication within 12 months after discharge from the hospital was 28.5% over this time period but varied by year. The treatment rates declined over the study period, from 2002 to 2011.2

The most striking correlate of post-fracture osteoporosis treatment was pre-fracture osteoporosis medication use. There was some variation in medication use by geographic U.S. Census Division; however, all regions demonstrated low osteoporosis treatment rates. 3

The testing data is reiterated here as an example of data demonstrating an opportunity for improvement. As recommended for inclusion here, the testing of this measure (see testing document) showed the following:
The Large Urology Group Practice Association (LUGPA) tested data from two datasets:

Dataset 1:
Information is from an operational electronic medical record from a large urology center in the West. We received an extract of EHR (GE Centricity Physicians EMR) data that included all patients meeting Initial Population (Denominator) and Numerator criteria for the measure. 100% manual analysis of the data was also done, which included a query within GE Centricity Physicians EMR as the initial query was from Practice Analytics that also included some patients that did not meet the initial population. To be included in the extracted data, the patient must be male with an office visit during 2017, have a prostate cancer diagnosis without bone metastasis, have osteopenia or osteoporosis and on prior or current androgen deprivation therapy.

Dataset 2:
Information is from an operational electronic medical record from a radiation center in the West that treats prostate cancer patients. We received a patient name extraction for all patients in 2017 from ARIA, a radiation center electronic medical record that is HL7 compliant and CMS approved. As you will see with the value set information, some of the information is not present in the radiation center EMR, but we were able to determine if the patient met the initial population and then if the patient met the numerator of the measure. We performed a 100% manual analysis of the data which had the initial query. To be included in the extracted data, the patient must be male with an office visit with the physician during 2017, have a prostate cancer diagnosis without bone metastasis, have osteopenia or osteoporosis and on prior or current androgen deprivation therapy.

The data is included from two (2) datasets. The data queried was for patients receiving care in the same group practice (also facility) for each dataset. Within the group practice, there were patients identified for each clinician and for Dataset 2, the patients were referred by multiple clinician partners, but being cared for by one clinician. A site (facility) was considered the element of analysis within each dataset. The below table shows the number of patients, clinicians and practices identified for each dataset.

	Patients	Clinicians	Sites
Dataset 1	65	11	1

Dataset 2 8 1 1

For Dataset 1, we generated reports from Practice Analytics and also from within GE Centricity Physicians EMR. The patient numbers were not large, so we chose to evaluate 100% of the reported patients. We initially evaluated for inclusion in the measure as any patient meeting the initial population. In looking at inclusion criteria, we evaluated not only the presence of the data element, but also compared the information within the data set to the inclusionary information within the value set as published by the Value Set Authority Center (VSAC).

Since we abstracted 100% of the patient charts, data included as "met" was in full agreement with the value sets. The below table shows the patients who met the measure versus the patients who met the initial population (patients reviewed). The data is separated by individual practitioners. To meet the measure, all data was required to be present. At the time of analysis, we reviewed the value sets which were present and verified their accuracy for this measure.

Physician	Patients Reviewed	Patients meet	Percentage meet
1	6	0	0.0%
2	4	3	75.0%
3	11	5	45.5%
4	2	0	0.0%
5	5	1	20.0%
6	4	0	0.0%
7	6	5	83.3%
8	2	1	50.0%
9	10	8	80.0%
10	8	7	87.5%
11	7	6	85.7%
Total	65	36	55.4%

Performance Rate range	0% to 87.5%
Average Performance Rate	47.91%
Number of Eligible Clinicians in the Data	11
Standard Deviation	35.27%

The specific value set and therefore data elements and agreement are included in the Feasibility Scorecard which is included within the Intent to Submit document.

For Dataset2, we generated a 100%patient list for 2017 report from ARIA. The patient numbers were also not large so we chose to evaluate 100% of the reported patients. As discussed earlier, this is a sub specialty EHR but is applicable to this measure as prostate cancer patients are being treated and often being placed on ADT during the treatment and for a period of 2 years post treatment. Some of these men have been on ADT prior. The focus on bone health has not been as aggressive as it has been at the prior site. This is not unusual and reflects the need for this measure and this focus. For those patients within this dataset, the value set inclusionary information was also evaluated. Where the patients in Dataset 1 are evaluated with a formal, inclusive Male Bone Health document that is not yet available in the Dataset 2 EHR. It is also not available in most EHR's, as the Dataset 1 process has been a national urology model for monitoring Male Bone Health for men on ADT.

The below information shows the results from the Dataset 2 manual extraction of 100% of patient.

Physician - ARIA EMR = 1
Patients reviewed - meeting initial population = 8
Patients meeting criteria for CMS 770 =0
% meet = 0%

From reference #2 below, The Osteoporosis Treatment Gap.
"None of these studies provide an insight into the causes underlying the substantial and increasing treatment gap. Factors that may play a role in the United States include a decline in BMD testing owing to reimbursement issues and lack of intensive

detailing by pharmaceutical companies. Solomon and colleagues point the finger at the lay press for raising awareness over the last decade of the potential side effects of the bisphosphonates, such as osteonecrosis of the jaw, atypical femoral fractures, and atrial fibrillation. Indeed, many doctors, dentists, and patients are now more frightened of the rare but serious side effects than they are of the disease and the fractures that arise. Notwithstanding, the lay press is simply the messenger bringing news and opinion from the scientific community, some or much of which may be ill-judged. The paradox arises in that we seek to treat individual patients to the highest standards but at the same time disservice and disadvantage the wider osteoporosis community. It is now time for us all to accept a long overdue collective responsibility for our failures and to work cohesively to improve the management of our patients. One hope is decreasing the treatment gap in the international development of fracture liaison services to better identify patients who have had a fragility fracture."

1International Osteoporosis Foundation. Gaps and Solutions in Bone Health: A Global Framework for Improvement, 2016. http://share.iofbonehealth.org/WOD/2016/thematic-report/2016TR-key-messages.pdf
2Kanis,J; Svedbon,A; Harvey,N; and McCloskey,E. Journal of Bone and Mineral Research. The Osteoporosis Treatment Gap, 2014. http://onlinelibrary.wiley.com/doi/10.1002/jbmr.2301/pdf

3Solomon, DH; Johnston,SS; Boytsov,NN; McMorrow,D; Lane,JM; Krohn,KD. Journal of Bone and Mineral Research. Osteoporosis medication use after hip fracture in US patients between 2002 and 2011, 2014. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4258070/

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*
N/A

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4
The main disparity in care for this condition lies within the difference in treatment between men and women. This was also summarized in 1.b.3. For many years, there has been recognition of osteoporosis in women, especially with the hormonal loss of menopause. The ADT causes secondary osteoporosis but has often been overlooked for treatment. This gender disparity is great and needs to be corrected as skeletal related events can lead to pain and disability as well as reduction in length of life for men with this condition.
Insurance disparities can also result in delays for treatment. Prostate cancer is a disease that often is found after the age of 50, although unfortunately younger men can acquire this cancer. Some insurance companies will not cover the necessary bone density screening (DEXA Scan) as a matter of the patient reaching a set age. Instead, the patient must have a high-risk drug (like ADT) or a fall with a fracture that is suspicious of decreased bone density. A proposed paired measure (CMS645) has a 3-month period between ADT start and DEXA scan requirement to meet the measure. This is due to the insurance disparity which requires the patient to be on a high-risk drug (ADT) in order to qualify for insurance coverage of a DEXA scan. As noted in the reference in 1.b.3, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4258070/ Osteoporosis Medication Use After Hip Fracture in U.S. Patients Between 2002 and 2011 (Solomon,D et al), J Bone Miner Res 2014, another possible cause of the low treatment rates may stem from the fragmented nature of the U.S. health care system. Communication from the in-hospital orthopedic surgeon after a hip fracture and the primary care provider can be difficult, resulting in less emphasis on osteoporosis care.
The U.S Census Bureau Division defined regions and the Kaplan-Meier estimated probabilities of osteoporosis medication use by regional geography resulted in 23.5% in the Middle Atlantic region to 30.2% in the east North Central region. Other regions were Pacific 30.1%, Mountain 27.6%, West North Central 29.0%, West South Central 27.6%, East South Central 27.8%, South Atlantic 28.5%, and New England 24.3%. Although the variance is not great between regions, the total rates are low throughout the United States, indicating a national performance gap.

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Treatment of osteopenia or osteoporosis in men with non-metastatic prostate cancer on androgen deprivation therapy
Date of Submission 5/18/2018
Type of Measure:

☐ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☒ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures**, section **2b3** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b5** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.

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

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **instrument-based measures** (including PRO-PMs) **and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **instrument-based measures (including PRO-PMs) and composite performance measures**, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]
AND
If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b3. For outcome measures and other measures when indicated (e.g., resource use):
- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR
- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

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

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

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions.
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure*. *Testing must be provided for all the sources of data specified and intended for*

measure implementation. ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ claims	☐ claims
☐ registry	☐ registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☒ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

The Large Urology Group Practice Association (LUGPA) tested data from two datasets:

Dataset 1:
Information is from an operational electronic medical record from a large urology center in the West. We received an extract of EHR (GE Centricity Physicians EMR) data that included all patients meeting Initial Population (Denominator) and Numerator criteria for the measure. 100% manual analysis of the data was also done, which included a query within GE Centricity Physicians EMR as the initial query was from Practice Analytics that also included some patients that did not meet the initial population. To be included in the extracted data, the patient must be male with an office visit during 2017, have a prostate cancer diagnosis without bone metastasis, have osteopenia or osteoporosis and on prior or current androgen deprivation therapy.

Dataset 2:
Information is from an operational electronic medical record from a radiation center in the West that treats prostate cancer patients. We received a patient name extraction for all patients in 2017 from ARIA, a radiation center electronic medical record that is HL7 compliant and CMS approved. As you will see with the value set information, some of the information is not present in the radiation center EMR, but we were able to determine if the patient met the initial population and then if the patient met the numerator of the measure. We performed a 100% manual analysis of the data which had the initial query. To be included in the extracted data, the patient must be male with an office visit with the physician during 2017, have a prostate cancer diagnosis without bone metastasis, have osteopenia or osteoporosis and on prior or current androgen deprivation therapy. All available data was extracted and all data elements were evaluated.

1.3. What are the dates of the data used in testing? Click here to enter date range
We chose 12 months of performance to be consistent with a one year measurement period within the measure specifications. We chose the most recent 12 months as the measure science is in use within these urology groups, but the exact process is being refined. Value sets were also tested during the extraction.

Dataset1: January 1, 2017 – December 31, 2017.

Dataset2: January 1, 2017 – December 31, 2017

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of:	**Measure Tested at Level of:**

(*must be consistent with levels entered in item S.20*)	
☒ individual clinician	☒ individual clinician
☒ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The data is included from two (2) datasets. The data queried was for patients receiving care in the same group practice (also facility)for each dataset. Within the group practice, there were patients identified for each clinician and for Dataset 2, the patients were referred by multiple clinician partners, but being cared for by one clinician. A site (facility) was considered the element of analysis within each dataset. The below table shows the number of patients, clinicians and practices identified for each dataset.

	Patients	Clinicians	Sites
Dataset 1	65	11	1
Dataset 2	8	1	1

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

The below table shows the number of patients, clinicians and sites identified for each dataset. Although Dataset 1 identifies one (1) site (facility), it includes 11 practices as all are within the same EHR and are partners in practice (group), thereby have the ability to follow similar processes.

	Patients	Clinicians	Sites
Dataset 1	65	11	1
Dataset 2	8	1	1

Patients in the dataset were based on the following inclusionary criteria (initial population):

- Male – no age limits, no race limits
- Diagnosis of Prostate Cancer (no bone metastasis)
- Diagnosis of Osteoporosis and/or Osteopenia – indicated as diagnosis on problem list. diagnosis on DXA scan, and/or stated diagnosis in past medical history
- Androgen Deprivation Therapy (ADT)current or prior
- Encounter performed during the review period - 2017

The initial population is also the denominator. The initial population includes male, prostate cancer diagnosis, osteopenia and/or osteoporosis diagnosis, prior or current medication of Androgen Deprivation Therapy and an encounter during the measurement period.

The prostate cancer diagnosis is specific to a neoplasm of the prostate without secondary metastasis to the bone. Metastasis to another body system could be included. Osteopenia and/or osteoporosis were either an established diagnosis, included from stated past medical history, or were determined by the DXA scan on record. The androgen deprivation therapy is available in many combinations, manufacturers, and a couple of routes of administration. It is primarily an injectable, depot medication, but is also available as an implant for longer medication distribution.

The numerator contains the active medication of Calcium and Vitamin D. The value set contains numerous combinations of Calcium and Vitamin D. The inclusion of these options was specific per the technical expert panel as dietary intake of Calcium and Vitamin D are also considerations by the practitioner and excess calcium is to avoided. The supplemental dosages are determined by the practitioner. The laboratory test is specific to total Calcium and total Vitamin D, 25 Hydroxy. The value set for bisphosphonates includes all of the specific dosages and manufacturers. The value set including denosumab includes only the dosage and administration interval for non-bone metastatic prostate cancer which is currently under the name of Prolia.

This was not a sample of patients, but instead a 100% analysis of patients that met the initial population.

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

The data used are the measure data elements. Both reliability and validity were tested as well as exclusions. This is not an outcome measure and therefore is not risk adjusted. This was a 100% analysis, not a sampling. When data elements met during 100% analysis, it was viewed to be certain the same linkage to data elements was present and correlated with the VSAC value sets. Dataset2 does not record osteoporosis or DXA scans for radiation therapy. We were able to obtain the information on each patient extracted and therefore have reliable initial population data.

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

The social risk factors were not analyzed as once a patient presents for care, these variables are not used to stratify care. The data evaluated was from clinics that provide care regardless of ability to pay. The clinics also provide language translation when needed for patients.

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

All data elements were empirically tested for accuracy and correctness as there was 100% abstraction and requirement to view the actual data as included within the value sets. Two people extracted data and coordinated results.

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

See section 2b2 for validity testing of data elements.

Discussion:
As stated above, the data elements were empirically tested for accuracy and correctness. Signal- to- noise ratio was used as false or irrelevant data needed to be identified. The signal-to-noise analysis was most evident in the selection of data elements as the VSAC has numerous combinations which contain the data needed but also contain irrelevant data that cannot count for the measure. For this reason, a number of the data elements were developed by the LUGPA Steward.

The steps for this process are:

Review each data element to first see that it is inclusive of all available and applicable codes for the data set. These can be codes from different systems (i.e. – SNOMEDCT, ICD10) and are grouping value sets when there are multiple code systems involved.

Review the linkage of reported data that was actual within the EHR's tested. Realizing we are looking at 2 EHR's, we also considered other possible combinations or entries which could be applicable in other EHR's. For example, Calcium total can be reported within a separate lab test or within metabolic panels.
Reviewed actual linkages or linkages that would be developed with the initiation of a clinic trying to meet this measure. Many EHR's develop lists or templates for the practitioner to check off that a condition or test was considered. This serves as a summary tool for the practitioner, but also as a documentation tool that consideration was given to a situation. For example, EHR's would most likely have a total calcium value reporting with a LOINC code and easily determined as a structured reporting within the EHR. Some EHR's actually have the value reporting as structured data which then also links for measure reporting to an acceptable coding system, like LOINC. This full tracking to the final coding system and data element was determined. However, a situation like dental carries or poor dentition would most likely not be reporting unless there was some kind of check off. Having a check system, if poor dentition is noted, it can then be checked and linked to the appropriate SNOMED CT code for this problem.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)
As you review the feasibility scorecard, you notice there are greater than 11 data elements in this measure. The scorecard will not accommodate for analysis greater than 11 data elements. This was reviewed with NQF staff and they acknowledged the scorecard has problems and needs additional formatting. That was not completed at the time of this submission, so we were told to submit what we had. It also does not link analysis for 2 EHR's being tested. Therefore, we broke down the data elements for testing in an additional feasibility scorecard solely for the purpose of obtaining the analysis, which was then transferred to the feasibility scorecard. You will see 2 additional tabs in the feasibility scorecard. One is for Analysis Data Elements 1-10 and the other is for Analysis Data Elements 11-20. The CMS supplemental data elements were not included as these automatically populate from CMS.

The summary of analysis for the first 10 was measure feasibility 92.5% and percent of data elements currently feasible within domain ranged from 90-100%. The lowest score for current elements was 2 and all scores in future elements were 3.

The summary of analysis for the second 10 was measure feasibility 77.5% and percentage of data elements currently feasible within domain ranged from 70-100%. The lowest score for current elements was 2 and all scores in future elements were 3.

Combining these 20 data elements, the measure feasibility is 85% and future feasibility is 100%.

One rater rated 2 trials of each sample and found a Cohen's Kappa of 1 for the critical data elements.

The statistical results from the signal-to-noise analysis were 100% at the finalization as this was the analysis and discrimination needed to be certain the data elements were correct. For data elements not yet linked in the EHR, the predicted signal-to-noise analysis was 100%, but actual could not be determined until a measure is fully in use and is undergoing measure maintenance.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i*.e., what do the results mean and what are the norms for the test conducted?*)

There is strong agreement demonstrated because the data elements are all from the VSAC. 16 of the 20 critical data elements are currently reported in at least one of the EHR's tested and directly align with the content with the value set or the direct reference code. The 4 that are not currently reported do align with the VSAC value set or direct reference code, but for future reporting can be directly linked which aligns with the !00% future feasibility

_____

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The statistical analysis was signal-to -noise analysis with the direct observation of each and every data element on each patient.

For Dataset 1, we generated reports from Practice Analytics and also from within GE Centricity Physicians EMR. The patient numbers were not large, so we chose to evaluate 100% of the reported patients. We initially evaluated for inclusion in the measure as any patient meeting the initial population. In looking at inclusion criteria, we evaluated not only the presence of the data element, but also compared the information within the data set to the inclusionary information within the value set as published by the Value Set Authority Center (VSAC).

Since we abstracted 100% of the patient charts, data included as "met" was in full agreement with the value sets. The below table shows the patients who met the measure versus the patients who met the initial population (patients reviewed). The data is separated by individual practitioners. To meet the measure, all data was required to be present. At the time of analysis, we reviewed the value sets which were present and verified their accuracy for this measure.

Physician	Patients Reviewed	Patients meet	Percentage meet
1	6	0	0.0%
2	4	3	75.0%
3	11	5	45.5%
4	2	0	0.0%
5	5	1	20.0%
6	4	0	0.0%

7	6	5	83.3%
8	2	1	50.0%
9	10	8	80.0%
10	8	7	87.5%
11	7	6	85.7%
Total	**65**	**36**	**55.4%**

Performance Rate range	**0% to 87.5%**
Average Performance Rate	**47.91%**
Number of Eligible Clinicians in the Data	**11**
Standard Deviation	**35.27%**

The specific value set and therefore data elements and agreement are included in the Feasibility Scorecard which is included within the Intent to Submit document.

For Dataset2, we generated a 100%patient list for 2017 report from ARIA. The patient numbers were also not large so we chose to evaluate 100% of the reported patients. As discussed earlier, this is a sub specialty EHR but is applicable to this measure as prostate cancer patients are being treated and often being placed on ADT during the treatment and for a period of 2 years post treatment. Some of these men have been on ADT prior. The focus on bone health has not been as aggressive as it has been at the prior site. This is not unusual and reflects the need for this measure and this focus. For those patients within this dataset, the value set inclusionary information was also evaluated. Where the patients in Dataset 1 are evaluated with a formal, inclusive Male Bone Health document that is not yet available in the Dataset 2 EHR. It is also not available in most EHR's, as the Dataset 1 process has been a national urology model for monitoring Male Bone Health for men on ADT.

The below table shows the results from the Dataset 2 manual extraction of 100% of patient.

Physician - ARIA EMR	**patients reviewed - meeting initial population**	**patients meeting criteria for CMS 770**	**% meet**
1	8	0	0%

In both of these EHR's, the initial data is attached to an observation term or similar code. When the data is extracted for MIPS reporting, there is a data source within the EHR that links the code from the EHR to the nationally accepted code of LOINC, ICD10, RXNorm, etc. At times, the code in the EHR is the same as the nationally accepted code (such as ICD10), and sometimes it requires a translation. With this in mind, when an obs term was linking the data, the data trail was then followed to be certain it can or does link to the nationally accepted code.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The validity testing showed 100% correlation with the value set chosen or created with the VSAC and the data included in the patient record. The important correlation for this measure which is new and not fully operational within an EHR is for the data elements and value sets to be directly correlating to the information being measured. This is the only way to define and measure the quality within the measure.

The data elements were all 3's or fully meeting and measurable currently within the EHR for 16 of 20 data elements. The sum of the perfect data elements was 12 points per data element. Four (4) data elements were not fully operational

now, but could be within 6 months, as stated earlier. Those 4 data elements have a sum or 9 points per data element. A t-test cannot be calculated at this time for this data as it is only one sample set, and also one analysis. Since we only have current data from one data extraction, we cannot determine a t-test until this measure is operational and we again extract data. At that time, we can compare the data per element and determine the t-test.

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)

We interpret that the results demonstrate a refined process and valid results. There is a strong correlation between the data collected and the information in the VSAC. The only variation from perfect correlation is the current ability to report 4 of the data elements in the EHR.

_____

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b3*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

The exclusions tested were:

Denominator Exclusions-

- Men with metastatic prostate cancer to the bone
- Terminally ill patients on hospice

Numerator Exclusions-

- Poor dentition
- Inflammation of the gums
- Dental procedure
- Awaiting dental clearance
- History of or planned radiation therapy to the jaw
- Prior osteonecrosis of the jaw
- Denosumab patients
    - Hypocalcemia until corrected
    - Known hypersensitivity
- Bisphosphonates
    - Hypocalcemia until corrected
    - Known hypersensitivity
    - Creatinine clearance below 35mL/mm
    - Barrett's esophagus

Denominator Exceptions-

- Medication order not done – Patient Reason refused

As part of the Intent to Submit process, this measure has been evaluated as a measure already in use. That is correct, but it is not in use to the extent that the defined measure will require. By this, we mean the patient is evaluated for all of the above situations and this is information included within the Male Bone Health document and within the EHR. However, it is information gathered at this time and is not currently fully linked to observation terms or codes within value sets. This is where the science of the measure has been carefully reviewed and delineated by the technical expert

panel and the measure is defining the linkages needed to accurately report the exclusion criteria within the EHR. Within the current EHR for the practice reviewed, the presence of the exclusion criteria makes it obvious to the practitioner to further evaluate, postpone care or decline offering a service due to critical issues needing addressed.

Exclusions do affect overall performance scores, but cannot be defined numerically as discussed above.

The Patient reason refused exception was tested within the CMS BONNIE testing. It has also been tested for correct linkage to the value set code system.

Since empirical data was used for all of the analysis, the statistical analysis was signal-to-noise. There was not extracted data that was not viewed and analyzed by human eyes.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

As stated in 2b2.1, it is not possible to measure the statistical results from testing exclusions and we are comfortable in stating it would be hard to find an EHR that tracks this information. This means this measure and the defined concerns in this measure are needed for proper monitoring of these patients and for proper reporting of male bone health and treatment of osteoporosis/osteopenia in men..

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusions are needed for two reasons. This first and most important reason is the exclusions indicate the practitioner conducted the evaluations and reviewed the information prior to offering osteoporosis/osteopenia treatment to these men. These exclusionary criteria aid in providing safe and quality care for these patients.

Secondly, the exclusions should be provided in order to reflect the true quality of care and not penalize the practitioner for not offering the care when it is clearly contraindicated.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

- ☐ **Published literature**
- ☐ **Internal data analysis**
- ☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The method used was the 100% extraction of patient data and therefore the statistical analysis was signal-to-noise. As shown in the below graph, there are practitioners who follow the guidelines closely and some who don't. Occasionally the only criteria missing were a record that the patient was taking calcium or vitamin D, or that the Vitamin D level was measured although the patient was on Vitamin D. These differences are statistically significant and meaningfully different. Although this does relate to the performance gap, it more directly relates to the more recent importance placed on male bone health for men on ADT.

Physician	Patients Reviewed	Patients meet	Percentage meet
1	6	0	0.0%
2	4	3	75.0%
3	11	5	45.5%
4	2	0	0.0%
5	5	1	20.0%
6	4	0	0.0%
7	6	5	83.3%
8	2	1	50.0%
9	10	8	80.0%
10	8	7	87.5%
11	7	6	85.7%
Total	**65**	**36**	**55.4%**

Performance Rate range	**0% to 87.5%**
Average Performance Rate	**47.91%**
Number of Eligible Clinicians in the Data	**11**
Standard Deviation	**35.27%**

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

The chart above indicates an average performance rate of 47.91% and a standard deviation of 35.27%. As noticed, some practitioners have 0% and some have a percentage nearly double the average. These are practitioners with the same resources available. It is known that clinically and practically speaking, some physicians place more emphasis than others on documenting male bone health. As we look at medical practice, there is still limited insurance and Medicare coverage for a DXA scan for a male, but much easier for a female to obtain a DXA scan. Yet, men placed on ADT are known to lose bone mineral density due to the medication. It is difficult for a physician to obtain the diagnostic tools when a patient is forced to pay out of pocket for a test that provides this critical information. Some patients decline testing and some greatly delay testing. In the last 10 years, denosumab has become available for men, both for prostate cancer with osteoporosis and for prostate cancer with bone metastasis. Bisphosphonates have been available for a longer period of time. It is known that skeletal related events are not only painful; they often shorten the life span for the patient and create a disability that greatly affects his quality of life.

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

Our interpretation of the results shows it is possible to measure the critical factors in this care. A e-clinical quality measure helps EHR using practitioners to track the critical values for care and to document the good care given, allowing them to receive credit within government programs. It also creates an incentive to follow the guidelines by proper reporting.

A recently accepted MIPS measure CMS645v1 is a precursor of this measure. It measures the patients with prostate cancer and placed on ADT who had a DXA scan within 3 months of the ADT order or initiation. This measure is the next step in performance for these men who have been shown to have osteoporosis or osteopenia.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
***If only one set of specifications, this section can be skipped*.**

This is an e-measure and has one set of specifications.

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

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

All records were directly reviewed and all data elements were directly correlated, therefore empirically evaluated which is the statistical analysis of signal-to-noise. Since this was not EHR reported and then manually extracted (measure is not operational for reporting in EHR), the agreement rate between the EHR and manual extraction cannot be determined. Missing data was only evident in 4 of the data sets, but when this information was missing, it was searched for in the narrative of the document. If the information was not in the narrative, then it was missing. However, this is not a measure in current use and, as stated, the information can be incorporated in an EHR in structured format and can be measured once this measure is fully operational in an EHR. For those 4 data sets that are not incorporated, they are defined in this measure and can easily be incorporated into this measure as has been discussed earlier.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

There was empirical sensitivity analysis and the information was determined in a narrative format when available. The main way for the missing data to be handled is through a template where the question is there for the practitioner to indicate if there is a concern, such as patient on comfort measures and thereby excluded from the measure. At times, this comment was in a narrative. There was not a demonstrated agreement between a EHR extraction and manual extraction as an EHR report for this measure is not available.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

There was empirical analysis. The results are clearly from current records. This measure is based upon sound science and the value sets are specific and applicable to each data element. Once the additional 4 data elements are fully reportable within an EHR, then this measure can be reported without bias.

2. Reliability and 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. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*

De.6. **Non-Condition Specific***(check all the areas that apply):*

De.7. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is an eMeasure **Attachment:** PCAwithosteoporosistreatment_v5_4_Artifacts_(35).zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Value_sets_for_Osteoporosis_with_ADT_measure04032018.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.
No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.
Not an instrument-based measure

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.
No

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Patients with an order for or taking bisphosphonates or denosumab and who had a Vitamin D and Calcium level completed prior to the start of treatment. Patients are also taking Calcium and Vitamin D.

S.5. **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, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator is the treatment of the patient with defined osteoporosis or osteopenia and must start in the measurement period. The osteoporosis or osteopenia is most commonly defined by a DEXA scan. (CMS645 is a MIPS measure defining DEXA scan at the time androgen deprivation therapy (ADT) is initiated). The patient must have an order for or be taking bisphosphonates or denosumab. Prior to the medication initiation, they must have a Vitamin D and Calcium level completed. They must also be taking Calcium and Vitamin D.
There are numerous numerator exclusions which must be addressed in this area as there is not a numerator exclusion area in the NQF documentation but there is one in the Measure Authoring Tool (MAT). Poor dentition, inflammation of the gums, dental procedure, awaiting dental clearance, and history of or planned radiation therapy to the jaw are all concerns for possible osteonecrosis of the jaw. Due to these concerns, dental clearance is required for the osteoporosis medications. Prior osteonecrosis of the jaw is an exclusion. For denosumab patients, hypocalcemia must be corrected prior to administration of the drug and known hypersensitivity to the medication is an exclusion. For bisphosphonate patients, hypocalcemia must also be correct, a creatinine clearance below 35 mL/min is an exclusion, they cannot have a history of Barrett's esophagus and known hypersensitivity to the medication is an exclusion. All included codes and value sets are listed in the excel document included in S2b (Value_sets_for_Osteoporosis_with_ADT_measure04032018)

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator equals the initial population. That is, male patients with a diagnosis of prostate cancer and osteoporosis or osteopenia. Patients with prior and/or current androgen deprivation therapy with an office visit during the measurement period. This is also the initial population.
There is no age cut off for this measure as prostate cancer can affect younger men, although it is a disease that normally occurs after the age of 40. According to the NCCN Prostate Cancer Early Detection guidelines, a cut off at 40 could miss those unfortunate patients who developed the disease in their late 20's and 30's. At the upper end, very healthy men over age 75 may choose to seek more aggressive treatment. Cancer genetics show an increased risk if the patient is a BRCA1/2 pathogenic mutation carrier which can lead to earlier detection of prostate cancers (and other cancers as well). When a family member is diagnosed with prostate cancer, another first degree relative is recommended to be screened at age 40 or 10 years prior to the age of the relative when prostate cancer was discovered, whichever is soonest.

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Male patient
Diagnosis of prostate cancer
Diagnosis of osteopenia or osteoporosis
Androgen Deprivation therapy starts before or prior to the end of the measurement period.
Office visit as included in the office visit value set during the measurement period.
All codes and values are listed in the excel spreadsheet in S2b (Value_sets_for_Osteoporosis_with_ADT_measure04032018).

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Denominator Exclusions are prostate cancer with secondary metastasis to the bone and patients on comfort measures such as hospice and end of life care.

The NQF document does not include a Denominator Exception, but there is one exception which is listed in the MAT - Patient refusal of the recommendation for bisphosphonates or denosumab after the start of ADT therapy and known osteoporosis diagnosis.

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
Prostate cancer with secondary metastasis to the bone is an ICD10 and SNOMED CT code in the patient chart. Comfort Measures including Hospice Care are included as SNOMED CT codes. Patient refusal of the medication must be in the measurement period and the EMR linkage to a SNOMED CT code. All of the codes and value sets are listed in S2b (Value_sets_for_Osteoporosis_with_ADT_measure04032018).

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*
Stratification is not required as this is not an outcome measure. It is a process measure.

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
No risk adjustment or risk stratification
If other:

S.12. **Type of score:**
Rate/proportion
If other:

S.13. **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

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or 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; time period for data, aggregating data; risk adjustment; etc.)*
The measure assesses performance on a health status and the associated process to address the problem. All of the measured evidence is obtainable through an electronic health record and directly refers to the scientific guidelines for treatment of osteoporosis or osteopenia in men with prostate cancer on androgen deprivation therapy. Following the evidence guide (Algorithm 1 of the Measure Evaluation criteria and Guidance for Evaluating Measures for Endorsement – August 2017), the evidence is rated as High.
The specific description of the measure calculation in order of sequence of steps are:
To identify the target population which is the initial population: Determine the sex of the patient and only include male patients. The patient must have a diagnosis of prostate cancer and a diagnosis of osteopenia or osteoporosis. The androgen deprivation therapy can start on or before the end of the measurement period. This allows for the patient diagnosed at the end of the prior year to have proper assessment prior to the start of the osteoporosis treatment. The initial population is also the denominator.
There are two exclusion which prohibit a patient from inclusion in the initial population. Men with metastatic prostate cancer to the bone are excluded as they would be on a different treatment protocol. Terminally ill patients, such as patients on hospice are also excluded.
A denominator exception is patient refusal of the recommended bisphosphonates or denosumab after the start of ADT therapy and a known osteoporosis diagnosis.
To identify the numerator: The patient must have an order for or are taking bisphosphonates or denosumab. They must have a Vitamin D and Calcium level prior to the start of treatment. They must also be taking supplemental Calcium and Vitamin D.
A patient would be excluded from the numerator if they have poor dentition, inflammation of the gums, if they have had a dental procedure, or if they are awaiting dental clearance. They are excluded if they have a history of or are planning to have radiation therapy to the jaw. Prior osteonecrosis of the jaw excludes them from treatment. For denosumab patients, they cannot start the medication if they have uncorrected hypocalcemia. They must not have a known hypersensitivity to the medication. For bisphosphonate patients, they also cannot start the medication if they have uncorrected hypocalcemia and/or known

hypersensitivity to the medication. They also cannot have been diagnosed with Barrett's Esophagus or have a creatinine clearance below 35mL/min.
The CQL version of this measure is attached. All value sets, timing and logic have been tested in the CMS BONNIE system. The testing was completed for the logic to pass and to fail, addressing all margins of the measure. Although the BONNIE testing excel output cannot be attached to this submission, it is being sent separately to the NQF staff as directed by them.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.
N/A

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
Specify calculation of response rates to be reported with performance measure results.
N/A

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Electronic Health Records

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*
IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.
The data is collected from an electronic health record, which includes CPT, SNOMEDCT,RXNORM, ICD10 and LOINC codes, as well as supplemental data elements (CDREC, SOP, Administrative Gender), all embedded within the electronic health record. The HL7 certified electronic health record reports the information in structured data for reporting.

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.20. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Individual

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Outpatient Services
If other:

S.22. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2. Validity – See attached Measure Testing Submission Form
nqf_testing_attachment_7.1_ 01052018.docx,copy of nqf_testing_attachment_7.1 04152018.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted

(prior testing as well as any new testing); use red font to indicate updated testing.

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

3. Feasibility

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

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

Most of the data sources are currently captured in the EHR´s tested. For those elements that are not currently captured in an EHR, they can be captured in an EHR. As discussed earlier and scored within the Feasibility Scorecard, there are several data elements that are not currently captured but are scored a 2. They are Comfort Measures, Dental indications surrounding administration of medications to treat osteoporosis in men, Oral radiation during medications for osteoporosis in men, and Patient Reason refused. All of these elements can be captured in a template with linkage to the data elements.

For comfort measures, there are 4 applicable SNOMED-CT codes. They are comfort measures, terminal care, dying care and hospice care which are all remine/therapy. If the practitioner indicates one of these therapies, then the SNOMED code will populate and report with this measure as a Denominator Exclusion of Terminally Ill Patients on Hospice. The only additional data entry from the provider is 1 click and should be designed to carry over to the next visit.

For Dental indications surrounding administration of medications to treat osteoporosis in men, these can also be indicated on a template. The advantage of a template is it allows the provider to review the necessary conditions and to quickly indicate a problem with 1 click. That information then populates the code system and also place a statement in the narrative of the EHR. If there are no dental problems, then the provider can mark dental clearance if the provider is capable of completing that exam. If they require a dentist to provide clearance, then the template can accept the dental clearance from the dentist. On subsequent

visits, the data can pull forward if the practice so designs it that way or it can be fields that need to be populated each visit. The dental indications value set contains 17 options which can be part of a drop down and can allow more than one to be checked. Some examples are sore gums, dental caries, poor dental hygiene, periodontal operation, swollen gums, drug-induced oral ulceration, and exposed bone in the jaw. These are SNOMED-CT and UMLS codes. These are part of the exam prior to further medication administration and are easily documented with a click. One of the clinics reviewed uses a Male Bone Health entry template to track information.
For Oral radiation during medications for osteoporosis in men, the indicator is one data element for the procedure of oral radiation. The design for a template would be to indicate recent or planned oral radiation. Both are considered Numerator Exclusions, but the modifier for the data element is different for each one. This is a SNOMED-CT code and can be accomplished with 1 click.
The last value set that is not currently captured is Patient Reason refused. This also is easily captured and is related to the osteoporosis medication. There are 6 options to indicate, all of which are situations and SNOMED-CT codes. Some examples are refusal of treatment by patient, patient non-compliant, refused intervention, and refused for religious reasons. That value carries forward if the patient refused and can be removed if the patient chooses to accept the recommendation in the future. This is a common negation rationale in current use in EHR's.

Comment for 3b.3 - The NQF Feasibility Score Card will accommodate 10 data elements for analysis of scoring, present and future. Attached is the copy sent to K Cobb on 01082018 and also an updated copy (05182018) with 2 additional tabs that break out the data elements for the measure (not supplemental data elements). The tabs are showing the analysis in 2 groups. NQF plans to modify this document to accommodate more data elements, but that is not yet available.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**
Attachment: Feasibility Scorecard_v1.0CMS770 for NQF GE and ARIA 04142018update05182018.xlsx

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None

4. Usability and Use

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

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Payment Program Regulatory and Accreditation Programs Quality Improvement (external benchmarking to organizations) Quality Improvement (Internal to the specific organization)	

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

N/A

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Developing this measures for CMS Merit-based Incentive Payment System (MIPS)

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

This measure is applicable to urology practices now and is a treatment that should be offered to patients when applicable. At one of the clinics where the patients were evaluated, the process is nearly 100% in place (missing easy report of the 4 date elements discussed earlier. Application of this process in its´ entirety could happen within 6 months. The clinic discussed has presented this process twice to other urology clinics through the Large Urology Group Practice Association. The specific plan is Male Bone Health Management with Osteoporosis Treatment if indicated. The purpose of the program is to provide proper monitoring and treatment for men undergoing continuous ADT, thereby having a higher risk for osteoporosis/osteopenia. The intended audience for the measure are urologists, but could include oncologists. As stated, the measure can be implemented within 6 months of acceptance. Most of the value sets are already linked in an EHR with diagnosis, medication and lab codes.

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

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

The process for evaluation and documentation of male bone health was presented in the draft form to the annual meeting of the Large Urology Group Practice Association (LUGPA) in 2015. This annual meeting was a west coast meeting and involved about 40 clinics. It was also presented to a regional meeting of LUGPA in 2015. It was discussed and specific elements reviewed. This meeting involved about 15-20 clinics. In fall of 2017, it was presented as a MIPS measure when discussing quality reporting at an

annual meeting of LUGPA. There were 2 break out sessions which were strongly attended by about 50 clinic representatives. This discussion was selected as the management of male bone health is an important discussion among the practitioners now that there are effective medications for this population. The sample was selected as all measures submitted by LUGPA and those submitted by Oregon Urology Institute were of interest to the general membership of LUGPA. The authorization of measures and development of the Technical Expert Panel occurred through the LUGPA Board of Directors.

4a2.1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

As above, the process has been presented 3 times since 2015 and most recently in the fall of 2017. Copies of the proposed QDM versions of the measure were provided when desired and a breakdown with discussion of all sections (numerator, denominator, etc) were reviewed. We opened it to discussion and questions from others as well. Some of the questions were simple, such as verification of value sets and other discussed the logic and science of the measure.

4a2.2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**

Describe how feedback was obtained.

The feedback on measure performance from entities that would be measured (practitioners and clinics) was positive. The measure had been refined by the LUGPA team, including the Technical Expert Panel (TEP) and the science of the measure was accurate. The discussion on the Oregon Urology Institute implementation of the measure was helpful for the attendees and some asked for screen shots of the actual form as well as the translation of the form to the narrative in the chart. Those who would be measured wanted to discuss implementation with their EHR vendor. The measure stewards also made themselves available for further questions and continue to do so.

This measure was submitted for 2017 Measures Under Consideration (MUC) and did not advance. We asked for an explanation and were advised additional testing needed to be provided as well as additional eCQM logic and BONNIE testing. The additional testing has been completed with advice and assistance from NQF. The 2017 MUC submission included a QDM version of the MAT. Since that time, the QDM version has been translated to CQL and the measure was tested in BONNIE to all margins and is 100% covered and 100% passing. Although the NQF submission document for this measure does not accommodate BONNIE testing, the excel format of the BONNIE testing is being submitted to NQF at their request and at the same time the measure is submitted.

4a2.2.2. **Summarize the feedback obtained from those being measured.**

LUGPA was favorable in having urology measures for reporting to MIPS as recently there were inadequate measures for urology specialty reporting. Also, with the break out sessions, we were able to have individual discussions where there were questions about certain data elements and we also had additional physician interest in participating with the Technical Expert Panel. The feedback has all been positive and any suggestions from these groups have been evaluated and/or implemented.

4a2.2.3. **Summarize the feedback obtained from other users**

Although not a user, CMS suggested further development and BONNIE testing which was done. CMS also asked us to consider pairing this with CMS645v1.

4a2.3. **Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**

When presented, the QDM version of the measure was discussed. At the same time we were discussing 4 other measures. We went through the initial population, denominator, denominator exclusions, numerator, numerator exclusions and denominator exceptions. They were most interested in the process by which we can delineate if the patient cannot meet the measure due to exclusions. These medications have very specific exclusions and we outlined the discussion and how the value sets work. Some of the participants were active in measure reporting and some were physicians who knew the high level view but did not know how to evaluate the data linkages within the EHR. The discussion was helpful for all as the attendees saw the level of concern with getting the measure correct and making it workable within an EHR. The measure was not modified as our technical expert panel and measure developers had been very inclusive of information. It was validating to have those questions asked and for the attendees to realize the material was considered.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
This measure and the information in 1b were developed over 2017 and to date 2018. This measure is not currently in use for performance improvement, but as stated, portions of the measure are used within Male Bone Health evaluation in clinics. The performance results could be used to further the goal of high-quality, efficient healthcare for both individuals and populations. Specifically, with the ability to report, most clinicians first reflect on what they are seeing on a month to month basis from EHR reports. As trending is being followed and measures for reporting are chosen, the clinician looks at their performance, what is lacking and what needs to be done to improve performance. Following quality improvement processes, one being the Plan-Do-Check-Act (PDCA) process assists in the clinic. In this situation, the initial plan is to review the measure and the linkages available within the EHR. Then, review what else is needed and how that can be documented and linked within the EHR. Training is involved once a template or documentation pattern is decided. It also involves full staff discussion to be certain all points are considered and who will be responsible. For example, a direct question about Calcium and Vitamin D supplementation during patient intake by the nurse or medical assistant will allow for the correct documentation to be entered and reviewable by the provider. The Do portion of the QI improvement involves full implementation of the documentation. This also involves the provider buy-in to the measure and importance of the measure. Male bone health has not been a focus over years and with the availability of the medication for usage in men with osteoporosis, the providers see an opportunity to reduce risk and maintain health in these men. The Check portion of the QI improvement is following the process and observing areas that are being omitted and also those areas that are consistently addressed. When there is an area of omission, the Act portion of the QI improvement is initiated as discussions occurs as to why that is happening and what can be done (and by whom) to be sure all elements are being considered. At this point, the cycle of PDCA is entered again to keep tabs on updates or revisions to determine the process is fully implemented and working.

4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**
This measure has not been fully implemented as it is a proposed measure.

4b2.2. **Please explain any unexpected benefits from implementation of this measure.**
No unexpected benefits have been found yet as this measure has not been fully implemented. It is a proposed measure.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications harmonized to the extent possible?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Only a partial target population is addressed with post fracture care in one measure. Please see discussion in 5b.1

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses 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.)
This measure does not address the same measure focus and the same target population as other NQF-endorsed measures. There are measures for treating women with osteoporosis (NQF0046), for screening for osteoporosis in women (MIPS 39), and osteoporosis management in women (NQF0053). NQF 0045 discusses communication with the physician or other clinician managing on-going care post fracture for men and women age 50 years and older. Although this has similarity for post fracture care, the measure being considered is for treatment of osteoporosis in men on ADT and who have prostate cancer which is known to attack bones.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Large Urology Group Practice Association
Co.2 **Point of Contact:** Colleen, Parker, colleen@oregonurology.com, 541-867-4474-
Co.3 **Measure Developer if different from Measure Steward:** Large Urology Group Practice Association
Co.4 **Point of Contact:** Colleen, Parker, colleen@oregonurology.com, 541-867-4474-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Bryan Mehlhaff, MD - Oregon Urology Institute and LUGPA - TEP member
Jeremy Shelton,MD - Skyline Urology and LUGPA - TEP member
Alec Koo, MD - Skyline Urology and LUGPA - TEP member
Robert Hollabaugh, MD - Conrad Pearson Clinic and LUGPA - TEP member
Paul Sieber, MD - Lancaster Urology and LUGPA - TEP member
Colleen Parker, RN, BSN, CEN - LUGPA and Oregon Urology Institute - measure steward and developer
Rachel Buchanan, MBA - LUGPA and Oregon Urology Institute - measure steward and developer

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?**
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:** Limited proprietary coding is contained in the measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. Oregon Urology Institute (OUI) and Large Urology Group Practice Association (LUGPA) disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT- [R]) or other coding contained in the specifications.

CPT (R) contained in the Measure specifications is copyright 2004-2017 American Medical Association. LOINC(R) copyright 2004-2017 Regenstrief Institute, Inc. This material contains SNOMED Clinical Terms(R) (SNOMED CT[R]) copyright 2004-2017 International Health Terminology Standards Development Organisation. ICD-10 copyright 2017 World Health Organization. All Rights Reserved.
Ad.7 **Disclaimers:** The performance measure is not a clinical guideline and does not establish a standard of medical care, and has not been tested for all potential applications.

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

Due to technical limitations, registered trademarks are indicated by (R) or [R].

Ad.8 **Additional Information/Comments:** NQF staff instructed us to send BONNIE testing excel documentation directly to staff. That is being done at the same time as submission.