



**To:** Consensus Standards Approval Committee (CSAC)  
**Fr:** Melissa Mariñelarena, Shaconna Gorham  
**Re:** Cancer 2015-2017  
**Da:** October 4, 2016

The CSAC will review recommendations from the Cancer project at its October 11 conference call.

This memo includes a summary of the project, recommended measures, and themes identified from and responses to the public and member comments.

Member voting on these recommended measures ended on September 21.

Accompanying this memo are the following documents:

1. [Cancer 2015-2017 Draft Report](#). The draft report has been updated to reflect the changes made following Standing Committee discussion of public and member comments. The complete draft report and supplemental materials are available on the project page.
2. [Comment table](#). Staff has identified themes within the comments received. This table lists 16 comments received and the NQF/Standing Committee responses.

## BACKGROUND

Cancer is the second most common cause of death in the U.S., exceeded only by heart disease.<sup>1</sup> The American Cancer Society estimated that almost 1.7 million new cases of cancer will be diagnosed in 2016. It also estimates approximately 314,000 men and 280,000 women will die from cancer in 2016<sup>2</sup>—that is more than 1,600 people a day. Furthermore, nearly half of all men and one-third of all women in the U.S. will develop cancer during their lifetime.<sup>3</sup> In addition to the loss of life, diagnosis and treatment of cancer has great economic impact on patients, their families, and society. In 2010, it was estimated that the costs for cancer care in the U.S. totaled nearly \$125 billion and could reach \$156 billion in 2020.<sup>4</sup>

The National Quality Forum's (NQF) portfolio of measures for cancer includes measures addressing cancer screening, appropriate treatment (including surgery, chemotherapy and radiation therapy), and morbidity and mortality.

For this project, the 24-member [Cancer Standing Committee](#) evaluated three newly-submitted measures and 15 measures undergoing maintenance of endorsement review against NQF's standard evaluation criteria. The Committee recommended 13 measures for endorsement, two measures for continued endorsement with reserve status, and three measures were not recommended. Evaluated measures are listed by topic in the draft report.

**DRAFT REPORT**

The Cancer 2015-2017 Draft Report presents the results of the evaluation of 18 measures considered under the Consensus Development Process (CDP). Thirteen measures are recommended for endorsement as voluntary consensus standards suitable for accountability and quality improvement, two for continued endorsement with reserve status, and three were not recommended for endorsement. The measures were evaluated against the 2015 version of the [measure evaluation criteria](#).

|   | <b>Maintenance</b>   | <b>New</b>   | <b>Total</b> |
|---|--|--|--------------|
| Measures under consideration                                      | 15   | 3  | 18           |
| Measures withdrawn from consideration                             | 1  | 3  | 4            |
| Measures recommended for endorsement                              | 11   | 2  | 13           |
| Measures recommended for inactive endorsement with reserve status | 2  | N/A  | 2            |
| Measures not recommended for endorsement                          | 2  | 1  | 3            |
| Reasons not Recommended   | Importance- 1<br>Scientific Acceptability- 1<br>Overall- 0<br>Competing Measure- 0 | Importance- 0<br>Scientific Acceptability- 1<br>Overall- 0<br>Competing Measure- 0 |              |

Pursuant to the CDP, the CSAC may consider approval of 13 candidate consensus measures and two measures recommended for inactive endorsement with reserve status.

**Cancer 2015-2017 Measures Recommended for Endorsement:**

- [0219](#): Post Breast Conservation Surgery Irradiation  
Overall Suitability for Endorsement: Y-19; N-1
- [0220](#): Adjuvant hormonal therapy  
Overall Suitability for Endorsement: Y-17; N-3
- [0223](#): Adjuvant Chemotherapy s Recommended or Administered within 4 Months (120 days) of Diagnosis to Patients Under the Age of 80 with AJCC III (lymph node positive) Colon Cancer  
Overall Suitability for Endorsement: Y-20; N-0
- [0225](#): At Least 12 Regional Lymph Nodes are Removed and Pathologically Examined for Resected Colon Cancer  
Overall Suitability for Endorsement: Y-17; N-3

- [0377](#): Hematology: Myelodysplastic Syndrome (MDS) and Acute Leukemias: Baseline Cytogenetic Testing Performed on Bone Marrow  
Overall Suitability for Endorsement: Y-21; N-0
- [0378](#): Hematology: Myelodysplastic Syndrome (MDS): Documentation of Iron Stores in Patients Receiving Erythropoietin Therapy  
Overall Suitability for Endorsement: Y-18; N-2
- [0389](#): Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients  
Overall Suitability for Endorsement: Y-19; N-0
- [2963](#): Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients - Legacy eMeasure  
Overall Suitability for Endorsement: Y-20; N-0
- [0390](#): Prostate Cancer: Adjuvant Hormonal Therapy for High or Very High Risk Prostate Cancer Patients  
Overall Suitability for Endorsement: Y-19; N-1
- [0508](#): Diagnostic Imaging: Inappropriate Use of “Probably Benign” Assessment Category in Screening Mammograms  
Overall Suitability for Endorsement: Y-20; N-1
- [0509](#): Diagnostic Imaging: Reminder System for Screening Mammograms  
Overall Suitability for Endorsement: Y-18; N-3
- [0559](#): Combination chemotherapy is recommended or administered within 4 months (120 days) of diagnosis for women under 70 with AJCC T1cN0M0, or Stage IB - III hormone receptor negative breast cancer  
Overall Suitability for Endorsement: Y-19; N-1
- [2930](#): Febrile Neutropenia Risk Assessment Prior to Chemotherapy  
Overall Suitability for Endorsement: Y-16; N-2

Cancer 2015-2017 Measures Recommended for Continued Endorsement with Reserve Status:

- [1878](#): HER2 Testing for Overexpression or Gene Amplification in Patients with Breast Cancer
- [1857](#): HER 2 Negative or Undocumented Breast Cancer Patients Spared Treatment with HER2-Targeted Therapies

Cancer 2015-2017 Measures Not Recommended (See [Appendix A](#) for the Committee’s votes and rationale):

- [0459](#): Risk-Adjusted Length of Stay >14 Days after Elective Lobectomy for Lung Cancer
- [0460](#): Risk-Adjusted Morbidity and Mortality for Esophagectomy for Cancer
- [2936](#): Admissions and Emergency Department (ED) Visits for Patients Receiving Outpatient Chemotherapy

## COMMENTS AND THEIR DISPOSITION

NQF received 15 comments from 3 member organizations and 2 members of the public (organizations and individuals) pertaining to the general draft report and to the measures under consideration.

A [table of comments](#) submitted during the comment period, with the responses to each comment and the actions taken by the Standing Committee and measure developers, is posted to the [Cancer 2015-2017 project page](#) under the Public and Member Comment section.

### **Comment Themes and Committee Responses**

Comments about specific measure specifications and rationale were forwarded to the developers, who were invited to respond.

The Standing Committee reviewed all of the submitted comments and developer responses. The Standing Committee focused their discussion on measures or topic areas with the most significant and recurring issues.

Three general themes were identified in the post-evaluation comments, including preference for outcome measures; request for changes to the measure description and specifications; and reserve status with inactive endorsement.

### ***Theme 1 – Preference for outcome measures***

Two measures, #2930 *Febrile Neutropenia Risk Assessment Prior to Chemotherapy* and #0378 *Hematology: Myelodysplastic Syndrome (MDS): Documentation of Iron Stores in Patients Receiving Erythropoietin Therapy* received two comments noting a preference for outcome measures. For #2930, the commenter noted that an outcome measure will assist in determining more than appropriate use of colony-stimulating factor (CSF), specifically resource utilization related to urgent care due to febrile neutropenia (FN). The commenter also noted the challenges of documenting FN risk assessment in electronic health records (EHR). For #0378, the commenter stated that it is unlikely that this measure will have a performance rate of 100%; therefore, an outcome measure based on the patient benefit of ESAs with respect to iron stores may be more appropriate.

*Developer Response (#2930):* Thank you for your comment. We agree that measuring febrile neutropenia (FN) outcomes is important, but view an outcome measure as a complement to our proposed measure rather than a substitute for two reasons.

First, the process measure is more actionable for oncology clinics which are our intended unit of analysis. With a process measure, clinics can set targets for improvement based on realistic expectations and relevant benchmarks and adopt management practices to reach those targets. On the other hand, an outcome measure would have to be risk-adjusted. So performance targets would have to be based on meaningful differences between expected and actual event rates, and performance could only be measured retrospectively. In addition, the ability to detect

performance differences depends on the sample size and the stability of the event rates. Both factors would make it difficult to use an FN outcome measure to inform management decisions at the clinic level.

Second, the intent of our proposed measure is to encourage appropriate use of CSF prophylaxis, i.e., promote use in patients with an elevated FN risk but discourage use in patients with a low risk, as defined by current guidelines. A standalone outcomes measure might incent clinics to overuse CSF prophylaxis to avoid adverse events and have the unintended consequence of overuse.

Thus, we believe that such a measure concept should be considered for future development, but it should not replace our proposed measure.

*Committee Response (#2930):* Thank you for your comment. The Committee agrees that a febrile neutropenia outcome measure would further the goal of high-quality, efficient healthcare rather than this process measure. However, the Committee also recognizes that certain process and structure measures are still useful for assessing quality, especially where outcomes may be difficult to measure. In addition, the Committee suggested incorporating the febrile neutropenia risk assessment into computerized physician order entry (CPOE) and standard orders to increase the feasibility of the measure in the future.

*Committee Response (#0378):* Thank you for your comment. The study conducted by Dr. Gregory Abel found that 56.0% of patients had evidence of pre-ESA iron assessment (Abel, G. A., Cronin, A. M., Odejide, O. O., Uno, H., Stone, R. M. and Steensma, D. P. (2016), Influence of patient and provider characteristics on quality of care for the myelodysplastic syndromes. *Br J Haematol*, 173: 713–721. doi:10.1111/bjh.13987).

The Committee agreed that this additional data suggests that a gap in performance exists in the documentation of iron stores in patients receiving erythropoietin therapy.

## ***Theme 2 – Request for changes***

A couple of comments suggested refining the measure description and specifications of two measures, #0559 *Combination Chemotherapy is Recommended or Administered Within 4 Months (120 Days) of Diagnosis for Women Under 70 with AJCC T1cN0M0, or Stage IB - III Hormone Receptor Negative Breast Cancer* and #0220 *Adjuvant Hormonal Therapy*.

*Developer Response (#0220 and #0559):* The American College of Surgeons, Commission on Cancer (CoC) thanks you for your comment and review of our measures. These quality measures use the terminology of administered within a specific timeframe or recommended based on the coding from the FORDS manual. This is the nationally standardized coding guideline promulgated by the CoC, and coordinated with several Federal agencies including the NCI and CDC; include specific code values indicating the clinical consideration of chemotherapy and hormone therapy and the choice of the patient and/or guardian to decline recommended therapy. Cancer registries within CoC-accredited cancer programs record and report this information if it is documented in the patient chart. The language of “recommended or

administered” in these measures was specifically selected after discussion with clinicians and users and is based directly on the FORDs data item definitions used to calculate these measures.

We agree with that when assessing overall quality, cancer programs should review patients in which treatment is administered and those in which treatment is recommended but not administered. Therefore, in the our reporting systems where compliance with these measures is assessed, cancer programs are able to view cases stratified by if; a) treatment is administered, b) treatment is recommended but not administered and c) the case is non-compliant with the measure. This allows programs to assess patients which cases are compliant with the measure but for which adjuvant therapy was not administered during internal quality improvement efforts.

For 2013 diagnoses cases in which treatment was recommended but not administered represents 6% of the numerator cases for measure #0220 changing compliance from 86.2% to 92.3%. For #0559, cases in which treatment was recommended but not administered represents 4% of the numerator changing overall compliance from 88.6% to 92.6%.

Another comment suggested measuring a different outcome for *#0459: Risk-Adjusted Length of Stay >14 Days after Elective Lobectomy for Lung Cancer*. The commenter suggested a measure addressing the discharge outcomes may provide better insight into variations of care due to low patient volume in the current measure. The commenter also noted the new measure(s) might be similar to measure #0460 with a different surgical procedure/patient diagnostic group.

*Developer Response (#0459):* STS appreciates the comment submitted by the Oncology Nursing Society. Although length of stay is a surrogate for morbidity, measure #0459 is intended to be used to measure health care resource utilization. STS serves as the measure developer and steward for NQF-endorsed measure #1790 Risk-Adjusted Morbidity and Mortality for Lung Resection for Lung Cancer, an outcomes measure that addresses the Oncology Nursing Society’s suggestion. In addition, STS recently developed a two-domain, outcomes only composite measure for lobectomy for lung cancer. The results of this composite have been distributed to STS General Thoracic Surgery Database participants, and planning is underway to add the lobectomy composite measure to STS’s voluntary public reporting program.

The Standing Committee reviewed the recommended changes, developer responses, and discussed these during the post-comment call before re-voting on the criteria where consensus was not reached for these measures.

### ***Theme 3 – Reserve status with inactive endorsement***

For the two measures that were recommended for reserve status with inactive endorsement, *#1857 HER2 Negative or Undocumented Breast Cancer Patients Spared Treatment with HER2-Targeted Therapies* and *#1878 HER2 Testing for Overexpression or Gene Amplification in Patients with Breast Cancer*, commenters requested changes to the specifications and/or a preference of outcome measures.

*Developer Response (#1857):* Thank you for your comments. The focus of this measure is to ensure patients receiving HER2 targeted therapies have documentation of a HER2 mutation.

ASCO agrees that determining whether the patient ever received HER2 testing is important and this aspect of care is addressed in NQF endorsed measure #1878 “HER2 testing for overexpression or gene amplification in patients with breast cancer.”

*Developer Response (#1857):* Thank you for your comment. ASCO recognizes the importance of outcome measures and efforts are in progress to develop these types of measures within the domains of oncology care.

*Developer Response (#1878):* Thank you for your response. ASCO continues to develop new measures and will consider developing a new measure to address disparities highlighted by this gap in practice.

*Developer Response (#1878):* Thank you for your comment. ASCO acknowledges that the data available are based on QOPI® self-selecting practices that voluntarily report data and may not be reflective of care provided outside of the QOPI® program.

*Committee Response:* Thank you for your comment. The Standing Committee will periodically review measures in reserve status for any change in evidence, evidence of deterioration in performance or unintended consequences, or any other concerns related to the measure. The Standing Committee may remove a measure from inactive endorsement status if the measure no longer meets NQF endorsement criteria. A maintenance review may occur upon a request from the Standing Committee or measure steward to return the measure to active endorsement.

#### **NQF MEMBER VOTING RESULTS**

All 15 of the recommended measures were approved with 80% approval or higher. Complete voting results are detailed in [Appendix B](#).

Representatives of 16 member organizations voted; no votes were received from the Public & Community Health Agency and Supplier/Industry Councils. (Links are provided to the full measure summary evaluation tables in [Appendix C](#).)

#### **REMOVE ENDORSEMENT OF MEASURES**

Five measures previously endorsed by NQF have not been re-submitted for maintenance of endorsement. One additional measure was withdrawn after the comment period.

| Measure   | Description  | Reason for removal of endorsement   |
|---|--|---|
| 0221: Image or Palpation-Guided Needle Biopsy (core or FNA) of the Primary Site is Performed to Establish Diagnosis of Breast Cancer (American College of Surgeons) | Percentage of patients presenting with AJCC Stage Group 0, I, II, or III disease, who undergo a needle biopsy to establish diagnosis of breast cancer. | Measure was not submitted for maintenance review. Developer determined they were not able conduct additional testing needed based on changes to measure specifications. |

| Measure  | Description   | Reason for removal of endorsement   |
|--|---|---|
| 0455: Recording of Clinical Stage Prior to Surgery for Lung Cancer or Esophageal Cancer Resection (The Society of Thoracic Surgeons) | Percentage of surgical patients aged 18 years and older undergoing resection for lung or esophageal cancer who had clinical staging provided prior to surgery   | Measure was not submitted for maintenance review. No reason provided by developer.  |
| 0457: Recording of Performance Status prior to Lung or Esophageal Cancer Resection (The Society of Thoracic Surgeons)                | Percentage of surgical patients aged 18 years and older undergoing resection for lung or esophageal cancer who had their performance status recorded within two weeks prior to the surgery date   | Measure was not submitted for maintenance review. No reason provided by developer.  |
| 0562: Overutilization of Imaging Studies in Melanoma (American Academy of Dermatology)   | Percentage of patients, regardless of age, with a current diagnosis of Stage 0 through IIC melanoma or a history of melanoma of any stage, without signs or symptoms suggesting systemic spread, seen for an office visit during the one-year measurement period, for whom no diagnostic imaging studies were ordered   | Measure was not submitted for maintenance review. The melanoma guideline is now in update and the developer anticipates developing new melanoma measures once the updated guideline is available. |
| 0650: Melanoma: Continuity of Care – Recall System (American Academy of Dermatology)   | Percentage of patients, regardless of age, with a current diagnosis of melanoma or a history of melanoma whose information was entered, at least once within a 12 month reporting period, into a recall system that includes: <ul style="list-style-type: none"> <li>• A target date for the next complete physical skin exam ,</li> <li>AND</li> <li>• A process to follow up with patients who either did not make an appointment within the specified timeframe or who missed a scheduled appointment</li> </ul> | Measure was not submitted for maintenance review. The melanoma guideline is now in update and the developer anticipates developing new melanoma measures once the updated guideline is available. |

## Appendix A – Measures Not Recommended for Endorsement

The table below lists the Committee's vote and rationale for measures not recommended for endorsement. Additional details are available via the number links.

| Measure  | Voting Results  | Rationale  |
|--|---|--|
| <b>0459: Risk-Adjusted Length of Stay &gt;14 Days after Elective Lobectomy for Lung Cancer</b> | <p><b>Initial Vote:</b><br/>Performance Gap<br/>H-0; M-10; L-5; I-6</p> <p><b>Post-Comment Call Vote:</b><br/>Performance Gap<br/>H-0; M-11; L-8; I-1</p>   | The Committee noted several concerns with the performance data provided by the developer and initially did not reach consensus on performance gap. During the in-person meeting, the Committee noted that the number of patients per region ranged from 2,996 per 40 surgeons to 7,756 patients per 73 surgeons, yet the mean prolonged length of stay (PLOS) was ~4.0% for each region. The Committee was concerned that low-volume providers may affect overall performance rates making it difficult to distinguish high-performers from low-performers and to determine if a gap in care exists based on the data provided. During the post-comment call the Committee discussed the same concerns. No additional data was provided and therefore the measure was not recommended for endorsement. |
| <b>0460: Risk-Adjusted Morbidity and Mortality for Esophagectomy for Cancer</b>                | <p><b>Initial Vote:</b><br/>Reliability:<br/>H-0; M-11; L-9; I-1<br/>Validity:<br/>M-12; L-9; I-0</p> <p><b>Post-Comment Call Vote:</b><br/>Reliability:<br/>H-2; M-9; L-8; I-1<br/>Validity:<br/>H-0; M-10; L-9; I-1</p> | The Standing Committee did not reach consensus on the reliability and validity criteria during the in-person meeting. The Standing Committee noted that more than 55.0% of participants (94) in the registry did fewer than five procedures a year. The Standing Committee expressed concern with the reliability of this low-volume procedure and that the measure was not specified for ≥5 procedures per year. The reliability of the measure score increased as the volume of minimum procedures per year for participants increased. The reliability scores for all 169 participants and 4,557 operations   |

| Measure   | Voting Results                            | Rationale   |
|---|---|---|
|   |   | <p>were 44.4%, 67.9% for <math>\geq 5</math> procedures per year, and 80.6% for <math>\geq 20</math> procedures per year.</p> <p>The Standing Committee also expressed concern with combining morbidity and mortality and asked the developer if there were plans for differential weighting of these outcomes. (The previous Committee also noted the same concern in 2012.) The developer stated that they were developing a new measure that more heavily weighs mortality than morbidity; measurement development is expected to be complete by the next maintenance review.</p> <p>The Committee determined that the data element validity testing provided was adequate but did not reach consensus on overall validity during the in-person meeting because low-volume providers was noted as a threat to validity. During the post-comment call the Committee discussed the same concerns. No additional data was provided and therefore the measure was not recommended for endorsement.</p> |
| <b>2936: Admissions and Emergency Department (ED) Visits for Patients Receiving Outpatient Chemotherapy</b> | <b>Reliability</b><br>H-0; M-4; L-10; I-6 | <p>The developer used 2012-2013 Medicare data from 3,765 hospitals and 240,446 patients. A total of 942 hospitals with <math>\geq 60</math> patients in the cohort were included in the sample. A split-sample methodology was used to test the measure score reliability. The reliability score for inpatient admissions was 0.41 and 0.27 for ED visits. During the workgroup call and the in-person meeting, the Committee questioned the developer about the strength of the reliability score for the ED measure (Pearson correlation = 0.27).</p> <p>In addition to their concerns with the</p>   |

| Measure | Voting Results | Rationale   |
|---------|----------------|---|
|         |                | <p>reliability scores of the measure, the Committee expressed their concerns with the numerator limiting admissions/rates to inpatient and ED. Many facilities and cancer centers have affiliated urgent care centers or 24-hour clinics rather than emergency departments. If a patient was seen at an urgent care centers or clinics for one of the eligible diagnoses, they would not be counted in the numerator. Additionally, if they were admitted to the hospital for observation, they would not be included in the numerator unless they crossed the two-midnight rule. Overall, the Committee concluded that the measure did not meet the reliability criterion due to the concerns discussed, specifically the small sample size used for reliability testing and the low reliability scores.</p> |

## Appendix B - NQF Member Voting Results

### Measure #0219 Post Breast Conservation Surgery Irradiation

| Measure Council                         | Yes       | No       | Abstain  | Total Votes | % Approval* |
|---|-----------|----------|----------|-------------|-------------|
| Consumer                                | 1         | 0        | 0        | 1           | 100%        |
| Health Plan                             | 0         | 0        | 1        | 1           |             |
| Health Professional                     | 6         | 0        | 2        | 8           | 100%        |
| Provider Organizations                  | 2         | 1        | 0        | 3           | 67%         |
| Public/Community Health Agency          | 0         | 0        | 0        | 0           |             |
| Purchaser                               | 2         | 0        | 0        | 2           | 100%        |
| QMRI                                    | 1         | 0        | 0        | 1           | 100%        |
| Supplier/Industry                       | 0         | 0        | 0        | 0           |             |
| <b>All Councils</b>                     | <b>12</b> | <b>1</b> | <b>3</b> | <b>16</b>   | <b>92%</b>  |
| Percentage of councils approving (>60%) |           |          |          |             | 100%        |
| Average council percentage approval     |           |          |          |             | 93%         |

\*equation: Yes/ (Total - Abstain)

#### Voting Comment:

- Adventist Health System: AHS believes that this measure has not met the NQF's reliability and validity criteria and therefore should not be eligible for NQF endorsement. In the report, the Committee notes that the developer submitted "updated mean performance rates" for the reliability evaluation. However, as stated in the report, "overall performance rates do not meet the reliability criterion." As such, the Committee chose to count data element validity testing for the data element reliability evaluation. However, "validity testing of all the critical data elements was not provided." AHS believes that the Committee clearly should have rated the reliability and validity as insufficient and decided not to recommend the measure until the developer has completed the necessary reliability and validity testing. The second question of the NQF's Measure Evaluation Criteria Algorithm 2. Guidance for Evaluating Reliability is: "Was empirical reliability testing conducted using statistical tests with the measures as specified?" We believe the Committee was correct in answering this question, "no." The next step, question 3 is: "Was empirical validity testing of patient-level data conducted?" AHS believes that the Committee incorrectly assessed this measure as meeting the criteria for validity. Question 2 of Algorithm 3. Guidance for Evaluating Validity asks, "Were all potential threats to validity that are relevant to the measure empirically assessed?" If, as the Committee has stated, validity testing for all the data elements, especially "critical" data elements, was not provided then the answer to this question should clearly be "no" and the measure should be rated as having insufficient validity and therefore also insufficient reliability.
- Although the Committee may feel that the measure specifications were consistently implemented, this does not satisfy the NQF's validity criteria. AHS finds that lax adherence to the NQF criteria, such as this, undermines the value of NQF endorsement. Unless reliability and validity testing is completed, especially for "critical" data elements, there is no way to verify that measures are indeed reliable and valid. Without this verifiability, NQF Endorsement has little meaning. It certainly does

not ensure that measures are valid and reliable. They might be or they might not be.

- AHS strongly disproves of endorsement of this measure at this time. We recommend that endorsement consideration of this measure be deferred until the developer completes reliability and validity testing. This should not be an issue because according to the report “the developer confirmed that they are planning to update their validity and reliability testing for the 5 measures submitted in this project (#0219, #0220, #0223, #0225 and #0559).”

#### Measure #0220 Adjuvant Hormonal Therapy

| Measure Council                         | Yes      | No       | Abstain  | Total Votes | % Approval* |
|---|----------|----------|----------|-------------|-------------|
| Consumer                                | 1        | 0        | 0        | 1           | 100%        |
| Health Plan                             | 0        | 0        | 1        | 1           |             |
| Health Professional                     | 5        | 0        | 3        | 8           | 100%        |
| Provider Organizations                  | 2        | 1        | 0        | 3           | 67%         |
| Public/Community Health Agency          | 0        | 0        | 0        | 0           |             |
| Purchaser                               | 0        | 1        | 1        | 2           | 0%          |
| QMRI                                    | 0        | 0        | 1        | 1           |             |
| Supplier/Industry                       | 0        | 0        | 0        | 0           |             |
| <b>All Councils</b>                     | <b>8</b> | <b>2</b> | <b>6</b> | <b>16</b>   | <b>80%</b>  |
| Percentage of councils approving (>60%) |          |          |          |             | 75%         |
| Average council percentage approval     |          |          |          |             | 67%         |

#### Voting Comment:

- Adventist Health System: AHS believes that this measure has not met the NQF’s reliability and validity criteria and therefore should not be eligible for NQF endorsement. In the report, the Committee notes that the developer submitted “updated mean performance rates” for the reliability evaluation. However, as stated in the report, “overall performance rates do not meet the reliability criterion.” As such, the Committee chose to count data element validity testing for the data element reliability evaluation. However, “validity testing of all the critical data elements was not provided.” AHS believes that the Committee clearly should have rated the reliability and validity as insufficient and decided not to recommend the measure until the developer has completed the necessary reliability and validity testing. The second question of the NQF’s Measure Evaluation Criteria Algorithm 2. Guidance for Evaluating Reliability is: “Was empirical reliability testing conducted using statistical tests with the measures as specified?” We believe the Committee was correct in answering this question, “no.” The next step, question 3 is: “Was empirical validity testing of patient-level data conducted?” AHS believes that the Committee incorrectly assessed this measure as meeting the criteria for validity. Question 2 of Algorithm 3. Guidance for Evaluating Validity asks, “Were all potential threats to validity that are relevant to the measure empirically assessed?” If, as the Committee has stated, validity testing for all the data elements, especially “critical” data elements, was not provided then the answer to this question should clearly be “no” and the measure should be rated as having insufficient validity and therefore also insufficient reliability.

- Although the Committee may feel that the measure specifications were consistently implemented, this does not satisfy the NQF's validity criteria. AHS finds that lax adherence to the NQF criteria, such as this, undermines the value of NQF endorsement. Unless reliability and validity testing is completed, especially for "critical" data elements, there is no way to verify that measures are indeed reliable and valid. Without this verifiability, NQF Endorsement has little meaning. It certainly does not ensure that measures are valid and reliable. They might be or they might not be.
- AHS strongly disapproves of endorsement of this measure at this time. We recommend that endorsement consideration of this measure be deferred until the developer completes reliability and validity testing. This should not be an issue because according to the report "the developer confirmed that they are planning to update their validity and reliability testing for the 5 measures submitted in this project (#0219, #0220, #0223, #0225 and #0559)."

**Measure #0223: Adjuvant Chemotherapy s Recommended or Administered within 4 Months (120 days) of Diagnosis to Patients Under the Age of 80 with AJCC III (lymph node positive) Colon Cancer**

| Measure Council                         | Yes       | No       | Abstain  | Total Votes | % Approval* |
|---|-----------|----------|----------|-------------|-------------|
| Consumer                                | 1         | 0        | 0        | 1           | 100%        |
| Health Plan                             | 1         | 0        | 0        | 1           | 100%        |
| Health Professional                     | 5         | 0        | 3        | 8           | 100%        |
| Provider Organizations                  | 2         | 1        | 0        | 3           | 67%         |
| Public/Community Health Agency          | 0         | 0        | 0        | 0           |             |
| Purchaser                               | 1         | 1        | 0        | 2           | 50%         |
| QMRI                                    | 0         | 0        | 1        | 1           |             |
| Supplier/Industry                       | 0         | 0        | 0        | 0           |             |
| <b>All Councils</b>                     | <b>10</b> | <b>2</b> | <b>4</b> | <b>16</b>   | <b>83%</b>  |
| Percentage of councils approving (>60%) |           |          |          |             | <b>80%</b>  |
| Average council percentage approval     |           |          |          |             | <b>83%</b>  |

## Voting Comment:

- Adventist Health System: AHS believes that this measure has not met the NQF's reliability and validity criteria and therefore should not be eligible for NQF endorsement. In the report, the Committee notes that the developer submitted "updated mean performance rates" for the reliability evaluation. However, as stated in the report, "overall performance rates do not meet the reliability criterion." As such, the Committee chose to count data element validity testing for the data element reliability evaluation. However, "validity testing of all the critical data elements was not provided." AHS believes that the Committee clearly should have rated the reliability and validity as insufficient and decided not to recommend the measure until the developer has completed the necessary reliability and validity testing. The second question of the NQF's Measure Evaluation Criteria Algorithm 2. Guidance for Evaluating Reliability is: "Was empirical reliability testing conducted using statistical tests with the measures as specified?" We believe the Committee was correct in answering this question, "no." The next step, question 3 is: "Was empirical validity testing of patient-level data conducted?" AHS believes that the Committee incorrectly assessed this measure as meeting the criteria for validity. Question 2 of Algorithm 3. Guidance for Evaluating Validity asks, "Were all potential threats to validity that are relevant to the measure empirically assessed?" If, as the Committee has stated, validity testing for all the data elements, especially "critical" data elements, was not provided then the answer to this question should clearly be "no" and the measure should be rated as having insufficient validity and therefore also insufficient reliability.
- Although the Committee may feel that the measure specifications were consistently implemented, this does not satisfy the NQF's validity criteria. AHS finds that lax adherence to the NQF criteria, such as this, undermines the value of NQF endorsement. Unless reliability and validity testing is completed, especially for "critical" data elements, there is no way to verify that measures are indeed reliable and valid. Without this verifiability, NQF Endorsement has little meaning. It certainly does not ensure that measures are valid and reliable. They might be or they might not be.
- AHS strongly disproves of endorsement of this measure at this time. We recommend that endorsement consideration of this measure be deferred until the developer completes reliability and validity testing. This should not be an issue because according to the report "the developer confirmed that they are planning to update their validity and reliability testing for the 5 measures submitted in this project (#0219, #0220, #0223, #0225 and #0559)."

**Measure #0225 At Least 12 Regional Lymph Nodes are Removed and Pathologically Examined for Resected Colon Cancer**

| Measure Council                | Yes | No | Abstain | Total Votes | % Approval* |
|--------------------------------|-----|----|---------|-------------|-------------|
| Consumer                       | 1   | 0  | 0       | 1           | 100%        |
| Health Plan                    | 0   | 0  | 1       | 1           |             |
| Health Professional            | 5   | 0  | 3       | 8           | 100%        |
| Provider Organizations         | 2   | 1  | 0       | 3           | 67%         |
| Public/Community Health Agency | 0   | 0  | 0       | 0           |             |
| Purchaser                      | 1   | 1  | 0       | 2           | 50%         |

|   |          |          |          |           |            |
|---|----------|----------|----------|-----------|------------|
| QMRI                                    | 0        | 0        | 1        | 1         |            |
| Supplier/Industry                       | 0        | 0        | 0        | 0         |            |
| <b>All Councils</b>                     | <b>9</b> | <b>2</b> | <b>5</b> | <b>16</b> | <b>82%</b> |
| Percentage of councils approving (>60%) |          |          |          |           | 75%        |
| Average council percentage approval     |          |          |          |           | 79%        |

\*equation: Yes/ (Total - Abstain)

#### Voting Comment:

- Adventist Health System: AHS believes that this measure has not met the NQF's reliability and validity criteria and therefore should not be eligible for NQF endorsement. In the report, the Committee notes that the developer submitted "updated mean performance rates" for the reliability evaluation. However, as stated in the report, "overall performance rates do not meet the reliability criterion." As such, the Committee chose to count data element validity testing for the data element reliability evaluation. However, "validity testing of all the critical data elements was not provided." AHS believes that the Committee clearly should have rated the reliability and validity as insufficient and decided not to recommend the measure until the developer has completed the necessary reliability and validity testing. The second question of the NQF's Measure Evaluation Criteria Algorithm 2. Guidance for Evaluating Reliability is: "Was empirical reliability testing conducted using statistical tests with the measures as specified?" We believe the Committee was correct in answering this question, "no." The next step, question 3 is: "Was empirical validity testing of patient-level data conducted?" AHS believes that the Committee incorrectly assessed this measure as meeting the criteria for validity. Question 2 of Algorithm 3. Guidance for Evaluating Validity asks, "Were all potential threats to validity that are relevant to the measure empirically assessed?" If, as the Committee has stated, validity testing for all the data elements, especially "critical" data elements, was not provided then the answer to this question should clearly be "no" and the measure should be rated as having insufficient validity and therefore also insufficient reliability.
- Although the Committee may feel that the measure specifications were consistently implemented, this does not satisfy the NQF's validity criteria. AHS finds that lax adherence to the NQF criteria, such as this, undermines the value of NQF endorsement. Unless reliability and validity testing is completed, especially for "critical" data elements, there is no way to verify that measures are indeed reliable and valid. Without this verifiability, NQF Endorsement has little meaning. It certainly does not ensure that measures are valid and reliable. They might be or they might not be.
- AHS strongly disproves of endorsement of this measure at this time. We recommend that endorsement consideration of this measure be deferred until the developer completes reliability and validity testing. This should not be an issue because according to the report "the developer confirmed that they are planning to update their validity and reliability testing for the 5 measures submitted in this project (#0219, #0220, #0223, #0225 and #0559)."

#### Measure #0377 Hematology: Myelodysplastic Syndrome (MDS) and Acute Leukemias: Baseline Cytogenetic Testing Performed on Bone Marrow

| Measure Council | Yes | No | Abstain | Total Votes | % Approval* |
|-----------------|-----|----|---------|-------------|-------------|
|-----------------|-----|----|---------|-------------|-------------|

|   |           |          |          |           |             |
|---|-----------|----------|----------|-----------|-------------|
| Consumer                                | 1         | 0        | 0        | 1         | 100%        |
| Health Plan                             | 0         | 0        | 1        | 1         |             |
| Health Professional                     | 6         | 0        | 2        | 8         | 100%        |
| Provider Organizations                  | 3         | 0        | 0        | 3         | 100%        |
| Public/Community Health Agency          | 0         | 0        | 0        | 0         |             |
| Purchaser                               | 2         | 0        | 0        | 2         | 100%        |
| QMRI                                    | 1         | 0        | 0        | 1         | 100%        |
| Supplier/Industry                       | 0         | 0        | 0        | 0         |             |
| <b>All Councils</b>                     | <b>13</b> | <b>0</b> | <b>3</b> | <b>16</b> | <b>100%</b> |
| Percentage of councils approving (>60%) |           |          |          | 100%      |             |
| Average council percentage approval     |           |          |          | 100%      |             |

\*equation: Yes/ (Total - Abstain)

**Measure #0378 Hematology: Myelodysplastic Syndrome (MDS): Documentation of Iron Stores in Patients Receiving Erythropoietin Therapy**

| Measure Council                         | Yes       | No       | Abstain  | Total Votes | % Approval* |
|---|-----------|----------|----------|-------------|-------------|
| Consumer                                | 1         | 0        | 0        | 1           | 100%        |
| Health Plan                             | 0         | 0        | 1        | 1           |             |
| Health Professional                     | 6         | 0        | 2        | 8           | 100%        |
| Provider Organizations                  | 3         | 0        | 0        | 3           | 100%        |
| Public/Community Health Agency          | 0         | 0        | 0        | 0           |             |
| Purchaser                               | 1         | 1        | 0        | 2           | 50%         |
| QMRI                                    | 1         | 0        | 0        | 1           | 100%        |
| Supplier/Industry                       | 0         | 0        | 0        | 0           |             |
| <b>All Councils</b>                     | <b>12</b> | <b>1</b> | <b>3</b> | <b>16</b>   | <b>92%</b>  |
| Percentage of councils approving (>60%) |           |          |          | 80%         |             |
| Average council percentage approval     |           |          |          | 90%         |             |

\*equation: Yes/ (Total - Abstain)

**Voting Comment:**

- Memorial Sloan-Kettering Cancer Center: If iron levels are normal and patient does not have bleeding or clinically apparent dietary changes or medical conditions which affect iron absorption, it should be considered acceptable to obtain iron (or ferritin – another measure of iron stores) levels more than 60 days prior to starting erythropoietin.

**Measure #0389 Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients**

| Measure Council     | Yes | No | Abstain | Total Votes | % Approval* |
|---------------------|-----|----|---------|-------------|-------------|
| Consumer            | 1   | 0  | 0       | 1           | 100%        |
| Health Plan         | 1   | 0  | 0       | 1           | 100%        |
| Health Professional | 7   | 0  | 1       | 8           | 100%        |

|   |           |          |          |           |             |
|---|-----------|----------|----------|-----------|-------------|
| Provider Organizations                  | 3         | 0        | 0        | 3         | 100%        |
| Public/Community Health Agency          | 0         | 0        | 0        | 0         |             |
| Purchaser                               | 2         | 0        | 0        | 2         | 100%        |
| QMRI                                    | 1         | 0        | 0        | 1         | 100%        |
| Supplier/Industry                       | 0         | 0        | 0        | 0         |             |
| <b>All Councils</b>                     | <b>15</b> | <b>0</b> | <b>1</b> | <b>16</b> | <b>100%</b> |
| Percentage of councils approving (>60%) |           |          |          | 100%      |             |
| Average council percentage approval     |           |          |          | 100%      |             |

\*equation: Yes/ (Total - Abstain)

**Measure #0390 Prostate Cancer: Adjuvant Hormonal Therapy for High or Very High Risk Prostate Cancer Patients**

| Measure Council                         | Yes       | No       | Abstain  | Total Votes | % Approval* |
|---|-----------|----------|----------|-------------|-------------|
| Consumer                                | 1         | 0        | 0        | 1           | 100%        |
| Health Plan                             | 0         | 0        | 1        | 1           |             |
| Health Professional                     | 7         | 0        | 1        | 8           | 100%        |
| Provider Organizations                  | 3         | 0        | 0        | 3           | 100%        |
| Public/Community Health Agency          | 0         | 0        | 0        | 0           |             |
| Purchaser                               | 2         | 0        | 0        | 2           | 100%        |
| QMRI                                    | 1         | 0        | 0        | 1           | 100%        |
| Supplier/Industry                       | 0         | 0        | 0        | 0           |             |
| <b>All Councils</b>                     | <b>14</b> | <b>0</b> | <b>2</b> | <b>16</b>   | <b>100%</b> |
| Percentage of councils approving (>60%) |           |          |          | 100%        |             |
| Average council percentage approval     |           |          |          | 100%        |             |

\*equation: Yes/ (Total - Abstain)

**Measure #0508 Diagnostic Imaging: Inappropriate Use of “Probably Benign” Assessment Category in Screening Mammograms**

| Measure Council                         | Yes       | No       | Abstain  | Total Votes | % Approval* |
|---|-----------|----------|----------|-------------|-------------|
| Consumer                                | 1         | 0        | 0        | 1           | 100%        |
| Health Plan                             | 0         | 0        | 1        | 1           |             |
| Health Professional                     | 5         | 0        | 3        | 8           | 100%        |
| Provider Organizations                  | 1         | 1        | 1        | 3           | 50%         |
| Public/Community Health Agency          | 0         | 0        | 0        | 0           |             |
| Purchaser                               | 2         | 0        | 0        | 2           | 100%        |
| QMRI                                    | 1         | 0        | 0        | 1           | 100%        |
| Supplier/Industry                       | 0         | 0        | 0        | 0           |             |
| <b>All Councils</b>                     | <b>10</b> | <b>1</b> | <b>5</b> | <b>16</b>   | <b>91%</b>  |
| Percentage of councils approving (>60%) |           |          |          | 80%         |             |
| Average council percentage approval     |           |          |          | 90%         |             |

\*equation: Yes/ (Total - Abstain)

Voting Comment:

- Adventist Health System: AHS is deeply concerned about this Committee's decision to override the NQF measure evaluation criteria. The Committee has noted that "based on NQF criteria, the evidence was insufficient due to lack of empirical evidence provided to support this process of care." However, despite this lack of evidence, the Committee decided to proceed with its recommendation based on its assessment that "it is beneficial to hold providers accountable for performance in the absence of empirical evidence of benefits to patients."
- AHS believes that endorsing measures that do not meet the NQF's criteria undermines the NQF's credibility and may erode consumer and provider confidence in quality measurement. We believe that it is possible that this process of care; "probably benign" should not be used as a category for indeterminate findings," could be beneficial for patients. Yet, we find that if NQF endorsement is to be the "gold standard" of quality measurement it is exceedingly important that the merits of such a process be supported by empirical evidence. If the NQF and its expert panels do not push measure developers to demonstrate the evidence behind their measures than there is little incentive to do so. This is clearly evident in this Committee's report. For example, in the evaluation of Measure 0225 it is noted that the previous Committee in 2012 had concerns with the quality of evidence in support of the measure. Yet, despite the developer's assurances that the measure would be "updated as the evidence evolved," the developer "did not provide updates to the evidence for the current [2016] endorsement evaluation." We believe that the reluctance of some NQF Committees to withhold recommendation until sufficient evidence is submitted stymies the advancement of quality measurement. It also derails efforts to identify "measures that matter."
- In order to achieve a more parsimonious and meaningful quality measure portfolio it is important to ensure that, in addition to identifying a performance gap, NQF-endorsed measures are evidence-based, reliable, valid, feasible and useful. Otherwise, health care providers must expend resources to report measures that may or may not reflect patient benefits and, more importantly, health care consumers are left sorting through a cacophony of measurement data that may or may not convey meaningful information regarding the potential benefits (or harms) of choosing a given provider or treatment.
- It is also important to note that the adherence to the NQF endorsement criteria varies widely across Committees. This Committee has chosen to recommend measures that lacked evidence or did not complete reliability testing or validity testing (Please see AHS' comments on Measures #0219, #0220, #0223, #0225 and #0559) Meanwhile, the Patient Safety Committee has strictly upheld the NQF evidence criteria in its recently released draft report. According to the report "the Committee did not find that sufficient evidence had been provided, so the measure [Measure #3005] was not recommended for endorsement." This decision was made even though the Committee recognized the importance of the measure and noted the existence of a performance gap.

**Measure #0509 Diagnostic Imaging: Reminder System for Screening Mammograms**

| Measure Council | Yes | No | Abstain | Total Votes | % Approval* |
|-----------------|-----|----|---------|-------------|-------------|
| Consumer        | 1   | 0  | 0       | 1           | 100%        |
| Health Plan     | 0   | 0  | 1       | 1           |             |

|   |          |          |          |           |            |
|---|----------|----------|----------|-----------|------------|
| Health Professional                     | 5        | 0        | 3        | 8         | 100%       |
| Provider Organizations                  | 1        | 1        | 1        | 3         | 50%        |
| Public/Community Health Agency          | 0        | 0        | 0        | 0         |            |
| Purchaser                               | 0        | 1        | 1        | 2         | 0%         |
| QMRI                                    | 1        | 0        | 0        | 1         | 100%       |
| Supplier/Industry                       | 0        | 0        | 0        | 0         |            |
| <b>All Councils</b>                     | <b>8</b> | <b>2</b> | <b>6</b> | <b>16</b> | <b>80%</b> |
| Percentage of councils approving (>60%) |          |          |          | 60%       |            |
| Average council percentage approval     |          |          |          | 70%       |            |

\*equation: Yes/ (Total - Abstain)

Voting Comment:

- Adventist Health System: We find that the usefulness of measure is limited since it does not access whether or not the reminder was actually sent.

**Measure #0559 Combination chemotherapy is recommended or administered within 4 months (120 days) of diagnosis for women under 70 with AJCC T1cN0M0, or Stage IB - III hormone receptor negative breast cancer**

| Measure Council                         | Yes      | No       | Abstain  | Total Votes | % Approval* |
|---|----------|----------|----------|-------------|-------------|
| Consumer                                | 1        | 0        | 0        | 1           | 100%        |
| Health Plan                             | 1        | 0        | 0        | 1           | 100%        |
| Health Professional                     | 5        | 0        | 3        | 8           | 100%        |
| Provider Organizations                  | 2        | 1        | 0        | 3           | 67%         |
| Public/Community Health Agency          | 0        | 0        | 0        | 0           |             |
| Purchaser                               | 0        | 1        | 1        | 2           | 0%          |
| QMRI                                    | 0        | 0        | 1        | 1           |             |
| Supplier/Industry                       | 0        | 0        | 0        | 0           |             |
| <b>All Councils</b>                     | <b>9</b> | <b>2</b> | <b>5</b> | <b>16</b>   | <b>82%</b>  |
| Percentage of councils approving (>60%) |          |          |          | 80%         |             |
| Average council percentage approval     |          |          |          | 73%         |             |

\*equation: Yes/ (Total - Abstain)

Voting Comment:

- Adventist Health System: AHS believes that this measure has not met the NQF's reliability and validity criteria and therefore should not be eligible for NQF endorsement. In the report, the Committee notes that the developer submitted "updated mean performance rates" for the reliability evaluation. However, as stated in the report, "overall performance rates do not meet the reliability criterion." As such, the Committee chose to count data element validity testing for the data element reliability evaluation. However, "validity testing of all the critical data elements was not provided." AHS believes that the Committee clearly should have rated the reliability and validity as insufficient and decided not to recommend the measure until the developer has completed the necessary reliability and validity testing. The second question of the NQF's Measure Evaluation Criteria Algorithm 2. Guidance for Evaluating Reliability is:

“Was empirical reliability testing conducted using statistical tests with the measures as specified?” We believe the Committee was correct in answering this question, “no.” The next step, question 3 is: “Was empirical validity testing of patient-level data conducted?” AHS believes that the Committee incorrectly assessed this measure as meeting the criteria for validity. Question 2 of Algorithm 3. Guidance for Evaluating Validity asks, “Were all potential threats to validity that are relevant to the measure empirically assessed?” If, as the Committee has stated, validity testing for all the data elements, especially “critical” data elements, was not provided then the answer to this question should clearly be “no” and the measure should be rated as having insufficient validity and therefore also insufficient reliability.

- Although the Committee may feel that the measure specifications were consistently implemented, this does not satisfy the NQF’s validity criteria. AHS finds that lax adherence to the NQF criteria, such as this, undermines the value of NQF endorsement. Unless reliability and validity testing is completed, especially for “critical” data elements, there is no way to verify that measures are indeed reliable and valid. Without this verifiability, NQF Endorsement has little meaning. It certainly does not ensure that measures are valid and reliable. They might be or they might not be.
- AHS strongly disapproves of endorsement of this measure at this time. We recommend that endorsement consideration of this measure be deferred until the developer completes reliability and validity testing. This should not be an issue because according to the report “the developer confirmed that they are planning to update their validity and reliability testing for the 5 measures submitted in this project (#0219, #0220, #0223, #0225 and #0559).”

**Measure #1857 HER 2 Negative or Undocumented Breast Cancer Patients Spared Treatment with HER2-Targeted Therapies – Recommended for Reserve Status**

| Measure Council                         | Yes       | No       | Abstain  | Total Votes | % Approval* |
|---|-----------|----------|----------|-------------|-------------|
| Consumer                                | 1         | 0        | 0        | 1           | 100%        |
| Health Plan                             | 1         | 0        | 0        | 1           | 100%        |
| Health Professional                     | 4         | 0        | 4        | 8           | 100%        |
| Provider Organizations                  | 3         | 0        | 0        | 3           | 100%        |
| Public/Community Health Agency          | 0         | 0        | 0        | 0           |             |
| Purchaser                               | 2         | 0        | 0        | 2           | 100%        |
| QMRI                                    | 0         | 0        | 1        | 1           |             |
| Supplier/Industry                       | 0         | 0        | 0        | 0           |             |
| <b>All Councils</b>                     | <b>11</b> | <b>0</b> | <b>5</b> | <b>16</b>   | <b>100%</b> |
| Percentage of councils approving (>60%) |           |          |          |             | 100%        |
| Average council percentage approval     |           |          |          |             | 100%        |

\*equation: Yes/ (Total - Abstain)

**Measure #1878 HER 2 Testing for Overexpression or Gene Amplification in Patients with Breast Cancer Recommended for Reserve Status**

| Measure Council | Yes | No | Abstain | Total Votes | % Approval* |
|-----------------|-----|----|---------|-------------|-------------|
| Consumer        | 1   | 0  | 0       | 1           | 100%        |

|   |           |          |          |           |             |
|---|-----------|----------|----------|-----------|-------------|
| Health Plan                             | 0         | 0        | 1        | 1         |             |
| Health Professional                     | 4         | 0        | 4        | 8         | 100%        |
| Provider Organizations                  | 3         | 0        | 0        | 3         | 100%        |
| Public/Community Health Agency          | 0         | 0        | 0        | 0         |             |
| Purchaser                               | 2         | 0        | 0        | 2         | 100%        |
| QMRI                                    | 0         | 0        | 1        | 1         |             |
| Supplier/Industry                       | 0         | 0        | 0        | 0         |             |
| <b>All Councils</b>                     | <b>10</b> | <b>0</b> | <b>6</b> | <b>16</b> | <b>100%</b> |
| Percentage of councils approving (>60%) |           |          |          | 100%      |             |
| Average council percentage approval     |           |          |          | 100%      |             |

\*equation: Yes/ (Total - Abstain)

#### Measure #2930 Febrile Neutropenia Risk Assessment Prior to Chemotherapy

| Measure Council                         | Yes      | No       | Abstain  | Total Votes | % Approval* |
|---|----------|----------|----------|-------------|-------------|
| Consumer                                | 1        | 0        | 0        | 1           | 100%        |
| Health Plan                             | 0        | 0        | 1        | 1           |             |
| Health Professional                     | 5        | 0        | 3        | 8           | 100%        |
| Provider Organizations                  | 2        | 1        | 0        | 3           | 67%         |
| Public/Community Health Agency          | 0        | 0        | 0        | 0           |             |
| Purchaser                               | 1        | 1        | 0        | 2           | 50%         |
| QMRI                                    | 0        | 0        | 1        | 1           |             |
| Supplier/Industry                       | 0        | 0        | 0        | 0           |             |
| <b>All Councils</b>                     | <b>9</b> | <b>2</b> | <b>5</b> | <b>16</b>   | <b>82%</b>  |
| Percentage of councils approving (>60%) |          |          |          | 75%         |             |
| Average council percentage approval     |          |          |          | 79%         |             |

\*equation: Yes/ (Total - Abstain)

#### Measure #2963

#### Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients - Legacy eMeasure

| Measure Council                         | Yes       | No       | Abstain  | Total Votes | % Approval* |
|---|-----------|----------|----------|-------------|-------------|
| Consumer                                | 1         | 0        | 0        | 1           | 100%        |
| Health Plan                             | 0         | 0        | 1        | 1           |             |
| Health Professional                     | 7         | 0        | 1        | 8           | 100%        |
| Provider Organizations                  | 1         | 1        | 1        | 3           | 50%         |
| Public/Community Health Agency          | 0         | 0        | 0        | 0           |             |
| Purchaser                               | 1         | 0        | 1        | 2           | 100%        |
| QMRI                                    | 1         | 0        | 0        | 1           | 100%        |
| Supplier/Industry                       | 0         | 0        | 0        | 0           |             |
| <b>All Councils</b>                     | <b>11</b> | <b>1</b> | <b>4</b> | <b>16</b>   | <b>92%</b>  |
| Percentage of councils approving (>60%) |           |          |          | 80%         |             |

|                                     |     |
|-------------------------------------|-----|
| Average council percentage approval | 90% |
|-------------------------------------|-----|

\*equation: Yes/ (Total - Abstain)

Voting Comment:

- Adventist Health System: This measure should not be endorsed until the developer has supplied the additional data necessary to fully evaluate the measure against the NQF endorsement criteria.

## Appendix C: Measure Evaluation Summary Tables

**LEGEND: Y = Yes; N = No; H = High; M = Moderate; L = Low; I = Insufficient**

### Measures Recommended

| 0219 Post breast conservation surgery irradiation  |
|--|
| <a href="#">Submission</a>   <a href="#">Specifications</a>  |
| <p><b>Description:</b> Percentage of female patients, age 18-69, who have their first diagnosis of breast cancer (epithelial malignancy), at AJCC stage I, II, or III, receiving breast conserving surgery who receive radiation therapy within 1 year (365 days) of diagnosis.</p> <p><b>Numerator Statement:</b> Radiation therapy to the breast is initiated within 1 year (365 days) of the date of diagnosis</p> <p><b>Denominator Statement:</b> Include, if all of the following characteristics are identified:<br/>         Women<br/>         Age 18-69 at time of diagnosis<br/>         Known or assumed to be first or only cancer diagnosis<br/>         Primary tumors of the breast<br/>         Epithelial malignancy only,<br/>         AJCC Stage I, II, or III<br/>         Surgical treatment</p> <p><b>Exclusions:</b> Exclude, if any of the following characteristics are identified:<br/>         Men<br/>         Under age 18 at time of diagnosis<br/>         Over age 69 at time of diagnosis<br/>         Second or subsequent cancer diagnosis<br/>         Tumor not originating in the breast<br/>         Non-epithelial malignancies<br/>         Phyllodes tumor histology<br/>         Stage 0, in-situ tumor<br/>         Stage IV, metastatic tumor<br/>         None of 1st course therapy performed at reporting facility<br/>         Died within 12 months (365 days) of diagnosis<br/>         Patient participating in clinical trial that directly impacts delivery of the standard of care</p> <p><b>Adjustment/Stratification:</b><br/> <b>Level of Analysis:</b> Facility<br/> <b>Setting of Care:</b> Hospital/Acute Care Facility<br/> <b>Type of Measure:</b> Process<br/> <b>Data Source:</b> Paper Medical Records, Electronic Clinical Data : Registry<br/> <b>Measure Steward:</b> Commission on Cancer, American College of Surgeons</p> |
| <p><b>STANDING COMMITTEE MEETING [5/18-19/2016]</b></p> <p><b>1. Importance to Measure and Report:</b> <u>The measure meets the Importance criteria</u><br/>         (1a. Evidence, 1b. Performance Gap)<br/>         1a. Evidence: <b>Previous Evidence Evaluation Accepted</b>; 1b. Performance Gap: <b>H-8; M-9; L-3; I-0</b>;</p>  |

**0219 Post breast conservation surgery irradiation**Rationale:

- For the 2012 endorsement evaluation, the developer provided a clinical practice guideline from National Comprehensive Cancer Network (NCCN) Practice Guidelines as evidence to support post breast conservation surgery irradiation. The developer also included a systematic review of multiple randomized clinical trials (RCTs) demonstrating a 75 percent reduction in the risk of local recurrence with radiation compared to no radiation in the hospital or acute care setting.
- The Committee agreed that the evidence basis for the measure has not changed and there was no need to repeat the discussion and vote on evidence.
- For the current evaluation, the developer provided national trend data from the National Cancer Data Base (NCDB) from 2008 and 2012. The mean performance rate for 2008 was 88.1% and 90.7% for 2012. The developer explained that more recent performance data was not available at the time of measure submission because participating programs have not yet had time to collect the 2013 information required for this measure.
- The Committee agreed that based on the performance and disparities data provided by the developer, a gap in care continues to exist for the utilization of radiation with breast conservation surgery for breast cancer.

**2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria**

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **M-16; L-2; I-2** 2b. Validity: **M-18; L-0; I-1**

Rationale:

- For the 2012 endorsement evaluation, the developer provided mean performance rates that included 1,400 Commission on Cancer (CoC) accredited cancer programs and approximately 55,700 cases from 2007 (84.1) and 2008 (84.7).
- For the current evaluation, the developer provided updated mean performance rates (90.7) from 2012. NQF reliability testing requirements include statistical analysis of the computed measure score or the individual patient-level data for the measured entities to determine the proportion of variation due to true differences versus noise or random variation. Overall performance rates do not meet the reliability criterion, which was provided by the developer. Data element validity testing was performed and counted for data element reliability.
- For the 2012 endorsement evaluation, validity was assessed by randomly selecting charts and reviewing them by site surveyors to determine completeness and validity of data reported to registry. The measure denominator and numerator were viewed by the clinical constituency within these cancer programs as valid and an appropriate reflection of the standard of care described in NCCN clinical guidelines.
- For the current evaluation, the developer provided additional details on data element validity testing conducted in 2009 and 2010 by comparing registry data to data that were re-abstracted from the medical records by CoC site surveyors, which was considered the gold standard. The developer provided percentage agreement results for one of the data elements included in the numerator (timing of radiation therapy (91.4, 92.2)). The developer also stated that there were 494 cases during 2006 and 2007 in which there was 59 percent agreement with missing radiation therapy. The Committee noted their concern with the large percentage of missing data required to calculate this measure as a threat to validity. The developer responded that they provide reports to the participating programs on missing data elements required to calculate the measures and verify whether or not the information is available. The developer also stated that they have seen a decrease in the percent of missing data since 2007 but did not provide updated testing information.
- One Committee member questioned why the measure only includes adults up to age 69. The developer responded that the age cutoff is based on the RCTs and that radiation therapy is generally considered necessary in women under age 69 versus a treatment preference in older

|  |
|--|
| <b>0219 Post breast conservation surgery irradiation</b>   |
| <p>women.</p> <ul style="list-style-type: none"> <li>Although validity testing of all the critical data elements (including kappa scores, sensitivity or specificity statistics) was not provided, the Committee agreed that the measure specifications were consistently implemented within the registry and met the reliability and validity criteria. The Committee encouraged the developer to provide updated reliability and validity testing at the next maintenance review of the measure. The developer confirmed that they are planning to update their validity and reliability testing for the 5 measures submitted in this project (#0219, #0220, #0223, #0225, and #0559).</li> </ul>                                  |
| <p><b>3. Feasibility: H-13; M-5; L-1; I-0</b></p> <p><i>(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)</i></p> <p><u>Rationale:</u></p> <ul style="list-style-type: none"> <li>The Committee agreed that, while a small level of burden exists, the measure is easily available in medical records and the data elements are routinely captured by national cancer registries.</li> </ul>   |
| <p><b>4. Usability and Use: H-12; M-6; L-1; I-0</b></p> <p><i>(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)</i></p> <p><u>Rationale:</u></p> <ul style="list-style-type: none"> <li>The measure is currently used in the Pennsylvania Health Care Quality Alliance, the Commission on Cancer, and the National Cancer Data Base reporting programs.</li> <li>The developer provided improvement results showing increases in the overall facility level compliance rates and across all patient demographics.</li> <li>The Committee agreed that the measure meets the usability and use criterion.</li> </ul> |
| <p><b>5. Related and Competing Measures</b></p> <ul style="list-style-type: none"> <li>No related or competing measures noted.</li> </ul>  |
| <b>Standing Committee Recommendation for Endorsement: Y-19; N-1</b>  |
| <p><b>6. Public and Member Comment</b></p> <ul style="list-style-type: none"> <li></li> </ul>  |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>   |
| <b>8. Board of Directors Vote: Y-X; N-X</b>  |
| <b>9. Appeals</b>  |

**0220 Adjuvant hormonal therapy**[Submission](#) | [Specifications](#)

**Description:** Percentage of female patients, age >18 at diagnosis, who have their first diagnosis of breast cancer (epithelial malignancy), at AJCC stage T1cN0M0,IB to III, who's primary tumor is progesterone or estrogen receptor positive with tamoxifen or third generation aromatase inhibitor (recommended or administered) within 1 year (365 days) of diagnosis.

**Numerator Statement:** Hormone therapy is administered within 1 year (365 days) of the date of diagnosis or it is recommended but not received

**Denominator Statement:** Include if all of the following characteristics are identified:

Women

Age >=18 at time of diagnosis

Known or assumed to be first or only cancer diagnosis

Epithelial malignancy only

Primary tumors of the breast

AJCC T1cN0M0 or Stage IB - III

Primary tumor

**Exclusions:** Exclude, if any of the following characteristics are identified:

Men

Under age 18 at time of diagnosis

Second or subsequent cancer diagnosis

Tumor not originating in the breast

Non-epithelial malignancies, exclude malignant phyllodes tumors, 8940 - Mixed tumor, malignant, NOS, 8950 - Mullerian mixed tumor, 8980 - Carcinosarcoma, 8981 - Carcinosarcoma, embryona

Stage 0, in-situ tumor

AJCC T1mic, or T1a tumor

Stage IV, metastatic tumor

Primary tumor is estrogen receptor negative and progesterone receptor negative

None of 1st course therapy performed at reporting facility

Died within 1 year (365 days) of diagnosis,

Patient enrolled in a clinical trial that directly impacts delivery of the standard of care

**Adjustment/Stratification:**

**Level of Analysis:** Facility

**Setting of Care:** Hospital/Acute Care Facility

**Type of Measure:** Process

**Data Source:** Paper Medical Records, Electronic Clinical Data : Registry

**Measure Steward:** Commission on Cancer, American College of Surgeons

**STANDING COMMITTEE MEETING [05/18-19/2016]****1. Importance to Measure and Report: The measure meets the Importance criteria**

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Previous Evidence Evaluation Accepted**; 1b. Performance Gap: **H-10; M-10; L-0; I-0**

**Rationale:**

- For the 2012 endorsement evaluation, the developer provided a clinical practice guideline from the National Comprehensive Cancer Network (NCCN) as evidence to support the administration of tamoxifen or third generation aromatase inhibitor to breast cancer patients whose primary tumor is progesterone or estrogen receptor positive. The developer also included results of a

**0220 Adjuvant hormonal therapy**

systematic review of several randomized control trials (RCTs) and meta-analyses demonstrating a 25% reduction in risk of distant cancer recurrence and death.

- The Committee agreed that the evidence basis for the measure has not changed and there was no need to repeat the discussion and vote on evidence.
- For the current evaluation, the developer provided national trend data from the National Cancer Data Base (NCDB) from 2008 and 2012. The mean performance rate for 2008 was 78.7% and 85.5% for 2012. The developer stated that the performance rate for 2013 was 90.1% (the most current data was not available at the time submission and will be submitted during the commenting period). The Committee noted that the performance data provided by the developer is from CoC-accredited centers only; therefore, the gaps in care and disparities may be larger if the measure was implemented in non-CoC-accredited centers.
- The Committee agreed that based on the performance and disparities data provided by the developer, a gap in care continues to exist in the administration of adjuvant hormonal therapy for breast cancer patients.

**2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability Criteria**

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

Initial 2a. Reliability: **M-9; L-9; I-2** 2b. Validity: **M-9; L-9; I-2**

Re-vote on 2a. Reliability: **H-2; M-12** 2b. Validity: **H-2; M-12**

**Rationale:**

- The developer clarified that the definition for “hormone therapy is administered” included documentation of a prescription and the date the prescription was filled or the date the treatment was started. The Committee noted that during the previous review of the measure, the previous Committee recommended that in future iterations, the measure capture that patients are receiving the appropriate dose of hormonal therapy, appropriateness of hormonal therapy based upon menopausal state of the patient, and patient adherence to the hormonal therapy through filled prescriptions. The developer did not update the measure specifications with any of these recommendations.
- For the 2012 endorsement evaluation, the developer provided mean performance rates that included 1,400 CoC-accredited cancer programs and approximately 65,200 cases from 2007 (76.6) and 2008 (77.1).
- For the current evaluation, the developer provided updated mean performance rates (85.5) from 2012. NQF reliability testing requirements include statistical analysis of the computed measure score or the individual patient-level data for the measured entities to determine the proportion of variation due to true differences vs. noise or random variation. Overall performance rates do not meet the reliability criterion, which was provided by the developer. Data element validity testing was performed and counted for data element reliability.
- For the 2012 endorsement evaluation, validity was assessed by randomly selecting charts and reviewing them by site surveyors to determine completeness and validity of data reported to registry. The measure denominator and numerator were viewed by the end-users within these cancer programs as valid and as an appropriate reflection of the standard of care described in NCCN clinical guidelines.
- For the current evaluation, the developer provided additional details on data element validity testing conducted in 2009 and 2010 by comparing registry data to data that were re-abstracted from the medical records by CoC site surveyors, which was considered the gold standard. The developer provided percentage agreement results for 2 of the data elements included in the numerator, timing for hormone therapy (84.3, 79.1) and hormone therapy which was recommended but not administered (77.9, 91.1). Validity testing of all the critical data elements (including kappa scores, sensitivity, or specificity statistics) was not provided. The Committee

|  |
|--|
| <b>0220 Adjuvant hormonal therapy</b>  |
| <p>noted that the measure specifications for this measure are not consistently implemented due to various patient factors such as the physician recommending hormone therapy, the patient obtaining a prescription, declining hormone therapy, and then possibly starting hormone therapy. The Committee also noted that the performance gap may not be accurate due to the variability in percent agreement of the data elements between the data submitted to the registry and the re-abstracted data. The Committee suggested this would lead to hospitals spending their resources to improve their performance rates on the measure rather than improving the overall quality of care for patients.</p> <ul style="list-style-type: none"> <li>The Committee encouraged the developer to provide updated reliability and validity testing at the next maintenance review of the measure. The developer confirmed that they are planning to update their validity and reliability testing for the 5 measures submitted in this project (#0219, #0220, #0223, #0225, and #0559).</li> </ul> |
| <p><b>3. Feasibility: H-1; M-15; L-4; I-0</b><br/> <i>(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)</i><br/> <u>Rationale:</u></p> <ul style="list-style-type: none"> <li>Although the data are readily available through medical records, the Committee recognized the data collection burden for manual chart abstraction that could result in various interpretations.</li> <li>The Committee agreed this measure meets the feasibility criterion.</li> </ul>   |
| <p><b>4. Usability and Use: H-10; M-6; L-4; I-0</b><br/> <i>(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)</i><br/> <u>Rationale:</u></p> <ul style="list-style-type: none"> <li>The measure is publicly reported through the Pennsylvania Health Care Quality Alliance, the PPS-Exempt Cancer Hospital Quality Reporting program, the Commission on Cancer, various compliance benchmarking programs through the National Cancer Data Base, and Quality Oncology Practice Initiative programs.</li> <li>The developer provided improvement results showing increases in the overall facility level compliance rates and across all census regions.</li> <li>The Committee agreed that despite some centers, including some of the PPS-exempt cancer hospitals, “topping out”, gaps persist among other hospitals, therefore, the performance results from this measure can continue to be used to further the goal of high-quality and efficient healthcare.</li> </ul>  |
| <p><b>5. Related and Competing Measures</b><br/> This measure is related to:</p> <ul style="list-style-type: none"> <li>#0387: Oncology: Hormonal therapy for stage IC through IIIC, ER/PR positive breast cancer (AMA-PCPI) <ul style="list-style-type: none"> <li>The developer stated that the measures are not harmonized because they assess different levels of analysis and different data systems are used to determine eligibility and compliance. Measure #0387 assesses whether hormone therapy was prescribed, whereas #0220 assesses whether hormone therapy was administered within one year of diagnosis or if it was recommended but not received based on patient refusal, medical co-morbidity, or other valid reasons. Measure #0220 assesses compliance at the facility level while #0387 assesses individual physician or practice level performance and the measures use different data sources as well.</li> </ul> </li> </ul>  |
| <p><b>Standing Committee Recommendation for Endorsement: Y-17; N-3</b></p>   |
| <p><b>6. Public and Member Comment</b></p>   |

|   |
|---|
| <b>0220 Adjuvant hormonal therapy</b>   |
| <ul style="list-style-type: none"> <li>• One commenter stated that it would be beneficial to have the measure stipulate administered vs. prescribed and to address who might not receive the treatment via the exclusions.</li> <li>• Developer response: The language of “recommended or administered” in these measures was specifically selected after discussion with clinicians and users and is based directly on the FORDs data item definitions used to calculate these measures. We agree with that when assessing overall quality, cancer programs should review patients in which treatment is administered and those in which treatment is recommended but not administered. Therefore, in the our reporting systems where compliance with these measures is assessed, cancer programs are able to view cases stratified by if; a) treatment is administered, b) treatment is recommended but not administered and c) the case is non-compliant with the measure. This allows programs to assess patients which cases are compliant with the measure but for which adjuvant therapy was not administered during internal quality improvement efforts.</li> <li>• During the Comment period, the developer submitted additional performance data from the Rapid Quality Reporting System (RQRS). The developer stated that the RQRS performance rates were similar to the performance rates from the NCDB.</li> <li>• The Committee considered the additional performance data from the Rapid Quality Reporting System (RQRS) and agreed this was an important indicator for cancer. On re-vote, the Committee recommended the measure for endorsement.</li> </ul> |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>  |
| <b>8. Board of Directors Vote: Y-X; N-X</b>   |
| <b>9. Appeals</b>   |

### 0223 Adjuvant chemotherapy is recommended or administered within 4 months (120 days) of diagnosis to patients under the age of 80 with AJCC III (lymph node positive) colon cancer

[Submission](#) | [Specifications](#)

**Description:** Percentage of patients under the age of 80 with AJCC III (lymph node positive) colon cancer for whom adjuvant chemotherapy is recommended and not received or administered within 4 months (120 days) of diagnosis.

**Numerator Statement:** Chemotherapy is administered within 4 months (120 days) of diagnosis or it is recommended and not received

**Denominator Statement:** Include, if all of the following characteristics are identified:

Age 18-79 at time of diagnosis

Known or assumed to be first or only cancer diagnosis

Primary tumors of the colon

Epithelial malignancy only

At least one pathologically examined regional lymph node

**Exclusions:** Exclude, if any of the following characteristics are identified:

Age <18 and ≥80; not a first or only cancer diagnosis; non-epithelial and non-invasive tumors; no regional lymph nodes pathologically examined; metastatic disease (AJCC Stage IV); not treated surgically; died within 4 months (120 days) of diagnosis; Patient participating in clinical trial which directly impacts receipt of standard of care.

**Adjustment/Stratification:**

**Level of Analysis:** Facility

**Setting of Care:** Hospital/Acute Care Facility

**Type of Measure:** Process

**Data Source:** Paper Medical Records, Electronic Clinical Data : Registry

**Measure Steward:** Commission on Cancer, American College of Surgeons

#### STANDING COMMITTEE MEETING [5/18-19/2016]

#### 1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Previous Evidence Evaluation Accepted**; 1b. Performance Gap: **H-6; M-14; L-0; I-0**

#### Rationale:

- For the 2012 endorsement evaluation, the developer provided a clinical practice guideline from the National Comprehensive Cancer Network (NCCN) recommending chemotherapy with Stage III colon cancer and a systematic review of the body of evidence demonstrating approximately 25% reduction in risk of death. The developer did not provide updates to the evidence for the current endorsement evaluation. The Committee agreed the evidence basis for the measure has not changed and accepted the previous evidence evaluation without further discussion.
- For the current evaluation, the developer provided national trend data from the National Cancer Data Base (NCDB) from 2008 and 2012. The mean performance rate for 2008 was 82.0% and 86.5% for 2012. The developer stated that the performance rate for 2014 was 86.2% (the most current data was not available at the time submission and will be submitted during the commenting period).
- A Committee member noted that the previous Committee questioned whether Stage 2b colon cancers should be included in the measure. At the time the developer responded that the evidence for the appropriateness of adjuvant chemotherapy for Stage 2b colon cancers was not complete. According to the developer, since the previous review, a German study concluded that Stage 2b colon cancers benefit slightly from adjuvant chemotherapy. The NCCN and American Society of Clinical Oncology (ASCO) guidelines recommend that adjuvant

**0223 Adjuvant chemotherapy is recommended or administered within 4 months (120 days) of diagnosis to patients under the age of 80 with AJCC III (lymph node positive) colon cancer**

chemotherapy is considered for Stage 2b colon cancers but the number in the study was considered insufficient for the recommendations for be implemented nationally.

- The Committee agreed the developer provided sufficient data on disparities based on race, ethnicity, age, insurance status, income, facility type, and sex and that a gap in care remains and there is opportunity for improvement.

**2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria**

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **M-13; L-6; I-1** 2b. Validity: **M-14; L-5; I-1**

Rationale:

- For the 2012 endorsement evaluation, the developer provided mean performance rates that included 1,400 CoC-accredited cancer programs and approximately 65,200 cases from 2007 (88.1) and 2008 (88.3).
- For the current evaluation, the developer provided updated mean performance rates (86.5) from 2012. NQF reliability testing requirements include statistical analysis of the computed measure score or the individual patient-level data for the measured entities to determine the proportion of variation due to true differences versus noise or random variation. Overall performance rates do not meet the reliability criterion, which was provided by the developer. Data element validity testing was performed and counted for data element reliability.
- For the 2012 endorsement evaluation, validity was assessed by randomly selecting charts and reviewing them by site surveyors to determine completeness and validity of data reported to registry. The measure numerator and denominator were viewed by the clinical constituency within these cancer programs as valid and an appropriate reflection of the standard of care described in NCCN clinical guidelines.
- For the current evaluation, the developer provided additional details on data element validity testing conducted in 2009 and 2010 by comparing registry data to data that were re-abstracted from the medical records by CoC site surveyors, which was considered the gold standard. The developer provided percentage agreement results for two of the data elements included in the numerator (timing of chemotherapy (88.9, 81.8) and therapy recommended but not received (88.5, 92.4)). Although validity of all the critical data elements (including kappa scores, sensitivity or specificity statistics) was not provided, the Committee agreed that the measure specifications were consistently implemented within the registry.
- The Committee agreed that the validity and reliability of the measure was sufficient but encouraged the developer to provide updated reliability and validity testing at the next maintenance review of the measure. The developer confirmed that they are planning to update their validity and reliability testing for the 5 measures submitted in this project (#0219, #0220, #0223, #0225, and #0559).

**3. Feasibility: H-8; M-11; L-1; I-0**

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee agreed that, while a small level of burden exists, the measure is easily available in medical records and the data elements are routinely captured by national cancer registries.

**4. Usability and Use: H-9; M-11; L-0; I-0**

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- The measure is currently used in the PPS-Exempt Cancer Hospital Quality Reporting program,

|  |
|--|
| <b>0223 Adjuvant chemotherapy is recommended or administered within 4 months (120 days) of diagnosis to patients under the age of 80 with AJCC III (lymph node positive) colon cancer</b>  |
| <p>Pennsylvania Health Care Quality Alliance, the Commission on Cancer, and the National Cancer Data Base reporting programs.</p> <ul style="list-style-type: none"> <li>• The developer provided improvement results showing increases in the overall facility level compliance rates and across all patient demographics.</li> <li>• The Committee agreed that the measure meets the usability and use criterion.</li> </ul>     |
| <p><b>5. Related and Competing Measures</b></p> <ul style="list-style-type: none"> <li>• This measure is related to: <ul style="list-style-type: none"> <li>○ #0385: Oncology: Chemotherapy for AJCC Stage III Colon Cancer Patients (AMA-PCPI)</li> </ul> </li> <li>• The measures assess different levels of analysis. #0223 assesses facility level performance; #0385 assesses clinical group practice performance.</li> </ul> |
| <b>Standing Committee Recommendation for Endorsement: Y-20; N-0</b>  |
| <p><b>6. Public and Member Comment</b></p> <ul style="list-style-type: none"> <li>•</li> </ul>   |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: 20-X; N-0</b>  |
| <b>8. Board of Directors Vote: Y-X; N-X</b>  |
| <b>9. Appeals</b>  |

## 0225 At least 12 regional lymph nodes are removed and pathologically examined for resected colon cancer.

[Submission](#) | [Specifications](#)

**Description:** Percentage of patients >18yrs of age, who have primary colon tumors (epithelial malignancies only), at AJCC stage I, II or III who have at least 12 regional lymph nodes removed and pathologically examined for resected colon cancer.

**Numerator Statement:** >=12 regional lymph nodes pathologically examined.

**Denominator Statement:** Include, if all of the following characteristics are identified:

Age >=18 at time of diagnosis

Primary tumors of the colon

Epithelial malignancy only

AJCC Stage I, II, or III

Surgical resection performed at the reporting facility

**Exclusions:** Exclude, if any of the following characteristics are identified: Age <18; non-epithelial and non-invasive tumors; metastatic disease (AJCC Stage IV); not treated surgically at the reporting facility; perforation of the primary site; acute obstruction

**Adjustment/Stratification:**

**Level of Analysis:** Facility

**Setting of Care:** Hospital/Acute Care Facility

**Type of Measure:** Process

**Data Source:** Paper Medical Records, Electronic Clinical Data : Registry

**Measure Steward:** Commission on Cancer, American College of Surgeons

### STANDING COMMITTEE MEETING [05/18-19/2016]

#### 1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **H-0; M-13; L-7; I-0**; 1b. Performance Gap: **H-6; M-12; L-2; I-0**

#### Rationale:

- For the 2012 endorsement evaluation, the developer provided a clinical practice guideline for stage II colon cancer from the National Comprehensive Cancer Network (NCCN). The guideline, based on lower-level evidence, states that if less than 12 lymph nodes are initially identified, it is recommended that the pathologist go back to the specimen and resubmits more tissue of potential lymph nodes. If 12 lymph nodes are still not identified, a comment in the report should indicate that an extensive search for lymph nodes was undertaken. The developer also provided a systematic review of the body of evidence that concluded that there is a lack of consensus on the minimal number of lymph nodes that have to be examined to accurately identify AJCC stage III colon cancer. The systematic review also concluded that an "adequate" lymph node examination was not associated with patient survival.
- The previous Committee in 2012 had noted their concern with the quality of the evidence presented and the lack of evidence demonstrating that 12 lymph nodes be identified. The developer stated that the measure would be updated as the evidence evolved.
- The developer did not provide updates to the evidence for the current endorsement evaluation. The Committee noted that the practice of examining 12 lymph nodes is not evidence-based, but rather an arbitrary number that is not connected to patient outcomes. During the Committee workgroup call, the developer stated that the National Cancer Data Base (NCDB) will be publishing a study demonstrating the relationship between compliance on this measure and outcomes over time. According to the recent studies conducted by NCDB, the developer stated, there is a correlation between the number of lymph nodes examined

**0225 At least 12 regional lymph nodes are removed and pathologically examined for resected colon cancer.**

and patient survival.

- Due to the low-level of evidence, the Committee decided to re-vote on the evidence criterion and agreed the evidence provided was sufficient at this time.
- For the current evaluation, the developer provided national trend data from the National Cancer Data Base (NCDB) from 2008 and 2013. The mean performance rate for 2008 was 81.7% and 89.7% for 2013. The Committee agreed that there was room for improvement though it is unlikely that the percentage will increase much more in high performers, since variation in surgical technique or pathology examination is likely to account for a significant number of patients that do not get to the 12 lymph node goal rather than poor performance.

**2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria**

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Previous Reliability Evaluation Accepted** 2b. Validity: **Previous Validity Evaluation Accepted**

Rationale:

- For the 2012 endorsement evaluation, the developer provided mean performance rates that included 1,400 CoC-accredited cancer programs and approximately 37,800 cases from 2007 (80.4) and 2008 (81.5).
- For the current evaluation, the developer provided updated mean performance rates (89.7) from 2012. NQF reliability testing requirements include statistical analysis of the computed measure score or the individual patient-level data for the measured entities to determine the proportion of variation due to true differences versus noise or random variation. Overall performance rates do not meet the reliability criterion, which was provided by the developer. Data element validity testing was performed and counted for data element reliability.
- For the 2012 endorsement evaluation, validity was assessed by randomly selecting charts and reviewing them by site surveyors to determine completeness and validity of data reported to registry. The measure denominator and numerator were viewed by the clinical constituency within these cancer programs as valid and an appropriate reflection of the standard of care described in NCCN clinical guidelines.
- For the current evaluation, the developer provided additional details on data element validity testing conducted in 2009 and 2010 by comparing registry data to data that were re-abstracted from the medical records by CoC site surveyors, which was considered the gold standard. The developer provided percentage agreement results for two of the data elements included in the numerator (timing of chemotherapy (88.9, 81.8) and therapy recommended but not received (88.5, 92.4)). Although validity of all the critical data elements (including kappa scores, sensitivity or specificity statistics) was not provided, the Committee agreed that the measure specifications were consistently implemented within the registry and accepted the previous reliability and validity evaluation.
- The Committee encouraged the developer to provide updated reliability and validity testing at the next maintenance review of the measure. The developer confirmed that they are planning to update their validity and reliability testing for the 5 measures submitted in this project (#0219, #0220, #0223, #0225, and #0559).

**3. Feasibility: H-6; M-11; L-3; I-0**

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- The Committee agreed that, while a small level of burden exists, the measure is easily available in medical records and the data elements are routinely captured by national cancer registries.

|   |
|---|
| <b>0225 At least 12 regional lymph nodes are removed and pathologically examined for resected colon cancer.</b>   |
| <p><b>4. Usability and Use: H-9; M-9; L-2; I-0</b><br/> <i>(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)</i></p> <p><u>Rationale:</u></p> <ul style="list-style-type: none"> <li>• The measure is currently used in the Pennsylvania Health Care Quality Alliance, the Commission on Cancer's accreditation program and National Cancer Data Base, and the Quality Oncology Practice Initiative (QOPI®).</li> <li>• The developer provided improvement results showing increases in the overall facility level compliance rates.</li> <li>• The Committee agreed that the measure meets the usability and use criterion.</li> </ul> |
| <p><b>5. Related and Competing Measures</b></p> <ul style="list-style-type: none"> <li>• No related or competing measures noted.</li> </ul>   |
| <b>Standing Committee Recommendation for Endorsement: Y-17; N-3</b>   |
| <p><b>6. Public and Member Comment</b></p> <ul style="list-style-type: none"> <li>•</li> </ul>  |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>  |
| <b>8. Board of Directors Vote: Y-X; N-X</b>   |
| <b>9. Appeals</b>   |

### 0377 Hematology: Myelodysplastic Syndrome (MDS) and Acute Leukemias: Baseline Cytogenetic Testing Performed on Bone Marrow

[Submission](#) | [Specifications](#)

**Description:** Percentage of patients aged 18 years and older with a diagnosis of myelodysplastic syndrome (MDS) or an acute leukemia who had baseline cytogenetic testing performed on bone marrow

**Numerator Statement:** Patients who had baseline cytogenetic testing performed on bone marrow

**Denominator Statement:** All patients aged 18 years and older with a diagnosis of myelodysplastic syndrome (MDS) or an acute leukemia

**Exclusions:** For Registry:

Documentation of medical reason(s) for not performing baseline cytogenetic testing (eg, no liquid bone marrow or fibrotic marrow)

Documentation of patient reason(s) for not performing baseline cytogenetic testing (eg, at time of diagnosis receiving palliative care or not receiving treatment as defined above)

Documentation of system reason(s) for not performing baseline cytogenetic testing (eg, patient previously treated by another physician at the time cytogenetic testing performed)

**Adjustment/Stratification:**

**Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Clinician : Team

**Setting of Care:** Ambulatory Care : Clinician Office/Clinic

**Type of Measure:** Process

**Data Source:** Electronic Clinical Data : Registry

**Measure Steward:** American Society of Hematology

#### STANDING COMMITTEE MEETING [05/18-19/2016]

##### 1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Previous Evidence Evaluation Accepted**; 1b. Performance Gap: **H-10; M-11; L-0; I-0**

##### Rationale:

- For the 2012 endorsement evaluation, the developer provided a lower-level evidence clinical practice guideline from the National Comprehensive Cancer Network (NCCN) for cytogenetic testing bone marrow of patients with myelodysplastic syndrome (MDS) and acute leukemia (AML).
- For the current evaluation, the Committee noted that the use of newer molecular cytogenetic studies using fluorescence in situ (FISH) is growing and encouraged the developer to include these additional studies. The developer agreed that as new evidence supporting additional studies continues to evolve and the guideline is revised, the measure will also be revised.
- The Committee agreed that higher-level evidence, such as randomized control trials (RCTs), supporting this measure would be limited; therefore, accepted prior evaluation of this criterion without further discussion.
- The developer provided average performance rates from the PQRS Registry from 2010 – 2013. The average performance rate was 88.8% in 2010, 94.6% in 2011, 95.6% in 2012, and 87.0% in 2013. The mean performance rate in 2014 was 95.09%, the minimum was 22.22%, and the maximum was 100.0%. The developer did not provide data on disparities from the measure as specified and stated they are not aware of any literature that addresses disparities in patients with ACL and MDS receiving baseline cytogenetic testing. The Committee agreed that performance has improved over time but there is still an opportunity for improvement.
- During the workgroup call, Dr. Gregory Abel stated that he was the primary author of a recently published study that demonstrated a 74% performance gap for this measure using Surveillance, Epidemiology, and End Results (SEER) – Medicare data. The developer will submit this additional

|   |
|---|
| <b>0377 Hematology: Myelodysplastic Syndrome (MDS) and Acute Leukemias: Baseline Cytogenetic Testing Performed on Bone Marrow</b>   |
| <p>information to NQF during the public commenting period.</p>  |
| <p><b>2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria</b><br/>         (2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)<br/>         2a. Reliability: <b>H-2; M-18; L-1; I-0</b> 2b. Validity: <b>M-20; L-1; I-0</b><br/> <u>Rationale:</u></p> <ul style="list-style-type: none"> <li>For the 2012 endorsement evaluation, inter-rater reliability was conducted on 29 acute leukemia patient records and 31 MDS patient records from 2 hematology practice sites. The percent agreement for the numerator was 98.3%, 100.0% for the denominator and the exclusions/exceptions, and 98.3% for overall reliability.</li> <li>For the current evaluation, the developer provided updated reliability testing of the measure score using a beta-binomial model to assess the signal-to-noise ratio. Reliability at the minimum level of quality reporting events (10) was 0.68 and 0.82 at the average number of quality events (21.0). A reliability of 0.70 is generally considered a minimum threshold for reliability.</li> <li>The Committee agreed that the updated reliability testing results were satisfactory and met the reliability criterion.</li> <li>For the 2012 endorsement evaluation, face validity of the measure score as an indicator of quality was systematically assessed by an expert panel. The expert panel agreed that the scores obtained from the measure as specified provide an accurate reflection of quality and can be used to distinguish good and poor quality.</li> <li>For the current evaluation, the developer conducted additional face validity testing with a panel of 23 experts representing the American Society of Hematology (ASH) Committee on Quality. Ninety-four percent of the total respondents (18) either agreed or strongly agreed that the measure can accurately distinguish good and poor quality. The Committee discussed the overall exclusion rate of 1.2% and determined that the exclusions are appropriate.</li> <li>The Committee agreed that the updated validity testing results were sufficient and met the validity criterion.</li> </ul> |
| <p><b>3. Feasibility: H-3; M-17; L-1; I-0</b><br/>         (3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)<br/> <u>Rationale:</u></p> <ul style="list-style-type: none"> <li>The Committee did not note any concerns regarding feasibility, acknowledging that the data elements used to construct this measure are based on clinical registry data and available in electronic sources.</li> </ul>   |
| <p><b>4. Usability and Use: H-4; M-17; L-0; I-0</b><br/>         (Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)<br/> <u>Rationale:</u></p> <ul style="list-style-type: none"> <li>The Committee noted that this measure is used in PQRS and will be available for public reporting on Physician Compare in late 2017.</li> </ul>   |
| <p><b>5. Related and Competing Measures</b></p> <ul style="list-style-type: none"> <li>No related or competing measures noted.</li> </ul>   |
| <p><b>Standing Committee Recommendation for Endorsement: Y-21; N-0</b></p>  |
| <p><b>6. Public and Member Comment</b></p> <ul style="list-style-type: none"> <li></li> </ul>   |
| <p><b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b></p>   |

|   |
|---|
| <b>0377 Hematology: Myelodysplastic Syndrome (MDS) and Acute Leukemias: Baseline Cytogenetic Testing Performed on Bone Marrow</b> |
| <b>8. Board of Directors Vote: Y-X; N-X</b>   |
| <b>9. Appeals</b>   |

### 0378 Hematology: Myelodysplastic Syndrome (MDS): Documentation of Iron Stores in Patients Receiving Erythropoietin Therapy

[Submission](#) | [Specifications](#)

**Description:** Percentage of patients aged 18 years and older with a diagnosis of myelodysplastic syndrome (MDS) who are receiving erythropoietin therapy with documentation of iron stores within 60 days prior to initiating erythropoietin therapy

**Numerator Statement:** Patients with documentation of iron stores within 60 days prior to initiating erythropoietin therapy

**Denominator Statement:** All patients aged 18 years and older with a diagnosis of myelodysplastic syndrome (MDS) who are receiving erythropoietin therapy

**Exclusions:** Documentation of system reason(s) for not documenting iron stores prior to initiating erythropoietin therapy

**Adjustment/Stratification:**

**Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Clinician : Team

**Setting of Care:** Ambulatory Care : Clinician Office/Clinic

**Type of Measure:** Process

**Data Source:** Electronic Clinical Data : Registry

**Measure Steward:** American Society of Hematology

#### STANDING COMMITTEE MEETING [05/18-19/2016]

##### 1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Previous Evidence Evaluation Accepted**; 1b. Performance Gap: **H-9; M-10; L-1; I-0**

##### Rationale:

- For the 2012 endorsement evaluation, the developer provided a lower-level evidence clinical practice guideline from the National Comprehensive Cancer Network (NCCN) that states that iron repletion be verified before instituting erythropoietin or darbepoetin therapy.
- For the current evaluation, the Committee agreed that higher-level evidence, such as randomized control trials (RCTs), supporting this measure would be limited; therefore, accepted prior evaluation of this criterion without further discussion.
- The developer provided average performance rates from the Physician Quality Reporting System (PQRS) Registry from 2010 – 2013. The average performance rate was 94.7% in 2010, 97.7% in 2011, 95.3% in 2012, and 83.1% in 2013. The mean performance rate in 2014 was 54.58%, the minimum was 0.0%, and the maximum was 100.0%. The developer did not provide data on disparities from the measure as specified and stated they are not aware of any literature outlining disparities for the documentation of iron stores in patients receiving erythropoietin therapy.
- The Committee agreed that performance has improved over time but there is still an opportunity for improvement.
- The recent study, conducted by Dr. Gregory Abel, and referenced during the [discussion for #0377](#), found that 56.0% of patients had pre-erythropoiesis-stimulating agent (ESA) assessments. The developer will submit this additional information to NQF during the public commenting period.

##### 2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-1; M-15; L-4; I-0** 2b. Validity: **M-16; L-3; I-1**

##### Rationale:

- For the 2012 endorsement evaluation, inter-rater reliability was conducted on 41

### 0378 Hematology: Myelodysplastic Syndrome (MDS): Documentation of Iron Stores in Patients Receiving Erythropoietin Therapy

myelodysplastic syndrome (MDS) patient records from 2 hematology practice sites. The percent agreement for the numerator was 90.2%, 100.0% for the denominator and the exclusions/exceptions, and 90.2% for overall reliability.

- For the current evaluation, the developer provided updated reliability testing of the measure score using a beta-binomial model to assess the signal-to-noise ratio. Reliability at the minimum level of quality reporting events (10) was 0.88 and 0.93 at the average number of quality events (18.4). A reliability of 0.70 is generally considered a minimum threshold for reliability.
- The developer clarified that to meet a portion of the denominator; the provider must attest that a qualifying patient is receiving erythropoietin therapy.
- The Committee encouraged the developer to consider including periodic monitoring of iron stores (in addition to baseline iron stores), to reduce the need for ESAs, maximize symptomatic improvement for patients, and determine the reason for failure to respond adequately to ESA therapy as currently indicated in the NCCN practice guideline.
- The Committee agreed that the updated reliability testing results were satisfactory and met the reliability criterion.
- For the 2012 endorsement evaluation, face validity of the measure score as an indicator of quality was systematically assessed by an expert panel. The expert panel agreed that the scores obtained from the measure as specified provide an accurate reflection of quality and can be used to distinguish good and poor quality.
- For the current evaluation, the developer conducted additional face validity testing with a panel of 23 experts representing the American Society of Hematology (ASH) Committee on Quality. Eighty-nine percent of the total respondents (18) either agreed or strongly agreed that the measure can accurately distinguish good and poor quality.
- The Committee questioned the developer about the seemingly excessive exclusion/exception rate of 97 exclusions/exceptions per 28 providers with an overall rate of 15.8%. The developer explained that they recommend providers document the specific reasons for exclusion/exception in patients' medical records for purposes of optimal patient management and audit-readiness. However, they are not able to obtain the specific reasons for not documenting iron stores prior to initiating erythropoietin therapy from PQRS data submitted to CMS. The developer also noted that, due to the high exclusion/exception rate, they have requested additional data from CMS to ensure that the measure is being reported accurately.
- Despite the potential threat to validity due to the high rate of exclusion/exception rates, the Committee agreed that without specific information about the exclusions/exceptions, it was difficult to understand how validity of the measure overall was impacted; therefore, the updated validity testing results were sufficient and met the validity criterion.

#### 3. Feasibility: H-5; M-13; L-1; I-1

*(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)*

##### Rationale:

- The Committee did not note any concerns regarding feasibility, acknowledging that the data elements used to construct this measure are based on clinical registry data and available in electronic sources.

#### 4. Usability and Use: H-7; M-12; L-1; I-0

*(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)*

##### Rationale:

- The Committee noted that this measure is used in PQRS and will be available for public reporting

|  |
|--|
| <b>0378 Hematology: Myelodysplastic Syndrome (MDS): Documentation of Iron Stores in Patients Receiving Erythropoietin Therapy</b>  |
| on Physician Compare in late 2017.   |
| <b>5. Related and Competing Measures</b> <ul style="list-style-type: none"> <li>No related or competing measures noted.</li> </ul>   |
| <b>Standing Committee Recommendation for Endorsement: Y-18; N-2</b>  |
| <b>6. Public and Member Comment</b> <ul style="list-style-type: none"> <li>One commenter stated that it is unlikely that this measure will have a performance rate of 100.0%; therefore, an outcome measure based on the patient benefit of ESAs with respect to iron stores may be more appropriate.</li> </ul> |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>   |
| <b>8. Board of Directors Vote: Y-X; N-X</b>  |
| <b>9. Appeals</b>  |

### 0389 Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients

[Submission](#) | [Specifications](#)

**Description:** Percentage of patients, regardless of age, with a diagnosis of prostate cancer at low (or very low) risk of recurrence receiving interstitial prostate brachytherapy, OR external beam radiotherapy to the prostate, OR radical prostatectomy, OR cryotherapy who did not have a bone scan performed at any time since diagnosis of prostate cancer

**Numerator Statement:** Patients who did not have a bone scan performed at any time since diagnosis of prostate cancer

**Denominator Statement:** All patients, regardless of age, with a diagnosis of prostate cancer at low (or very low) risk of recurrence, receiving interstitial prostate brachytherapy, OR external beam radiotherapy to the prostate, OR radical prostatectomy, OR cryotherapy

**Exclusions:** Documentation of medical reason(s) for having a bone scan performed (including documented pain, salvage therapy, other medical reasons)

Documentation of system reason(s) for having a bone scan performed (including bone scan ordered by someone other than reporting physician)

**Adjustment/Stratification:**

**Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Clinician : Team

**Setting of Care:** Ambulatory Care : Ambulatory Surgery Center (ASC), Ambulatory Care : Clinician Office/Clinic, Other

**Type of Measure:** Process

**Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry

**Measure Steward:** PCPI

#### STANDING COMMITTEE MEETING [05/18-19/2016]

##### 1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Previous Evidence Evaluation Accepted**; 1b. Performance Gap: **H-2; M-15; L-2; I-0**

##### Rationale:

- For the 2012 endorsement evaluation, the developer provided a best practice statement, a clinical practice guideline, and a systematic review of the body of evidence to demonstrate the use of bone scans for low risk prostate cancer patients is not supported by the evidence, is extremely costly, and unnecessarily exposes patients to radiation.
- For the current evaluation, the developer updated the evidence with updates to the best practice statement and clinical practice guideline. There were no changes to the recommendations since the previous submission.
- The developer also provided new evidence from the 2012 American College of Radiology (ACR) Appropriateness Criteria: Prostate Cancer – Pretreatment Detection, Staging, and Surveillance. The ACR criteria recommends that only patients with a PSA  $\geq 20$  ng/ml (with any T stage or Gleason score), locally advanced disease (T3 or T4 with any PSA or Gleason score), or Gleason score  $\geq 8$  (with any PSA or T stage) should be considered for a radionuclide bone scan. Patients with skeletal symptoms or advanced-stage disease should also be considered candidates for bone scans.
- The Committee agreed that the updated evidence supports the measure focus and has a stronger level of evidence. The Committee accepted the prior evaluation of this criterion without further discussion.
- The developer provided group/practice level performance data from 2014 PQRS EHR, Registry, and Part B Claims. The mean performance rate for EHR data was 90.76% and 90.24% for registry

### 0389 Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients

data. The developer also provided average performance rates from the PQRS Experience Report from 2010-2013. The average performance rate was 71.6% in 2010, 90.5% in 2011, 92.5% in 2012, and 88.5% in 2013. The developer did not provide data on disparities from the measure as specified but cited literature showing higher morbidity and mortality of prostate cancer in African-Americans. Another citation suggests that imaging overuse is associated with nonwhite race, education, income, and region. The Committee agreed that performance has improved over time but there is still an opportunity for improvement.

#### **2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria**

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Previous Reliability Evaluation Accepted** 2b. Validity: **Previous Validity Evaluation Accepted**

##### Rationale:

- For the 2012 endorsement evaluation, inter-rater reliability was conducted on 94 patient records from 2010; chart and data auditing occurred in 2011. The percent agreement for the numerator, denominator, exclusions/exceptions, and overall was 100.0%.
- For the current evaluation, the developer provided updated reliability testing of the measure score using a beta-binomial model to assess the signal-to-noise ratio. Reliability at the minimum level of quality reporting events (10) was 0.84 and 0.96 at the average number of quality events (46.0).
- The Committee agreed that the updated reliability testing results were satisfactory and accepted the prior evaluation of this criterion without further discussion.
- For the 2012 endorsement evaluation, face validity of the measure score as an indicator of quality was systematically assessed by an expert panel. The expert panel agreed that the scores obtained from the measure as specified provide an accurate reflection of quality and can be used to distinguish good and poor quality.
- For the current evaluation, the developer conducted additional face validity testing with a panel of 17 experts representing the PCPI Measures Advisory Committee. A total of 80% (10) of the respondents either agreed or strongly agreed that the measure can accurately distinguish good and poor quality. The Committee discussed the overall exclusion rate of 14.1% and determined that the exclusions are appropriate.
- The Committee agreed that the updated validity testing results were sufficient and accepted the prior evaluation of this criterion without further discussion.

#### **3. Feasibility: H-10; M-9; L-0; I-0**

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)

##### Rationale:

- The Committee did not note any concerns regarding feasibility, acknowledging that the data elements used to construct this measure are based on clinical registry data and available in electronic sources.

#### **4. Usability and Use: H-13; M-6; L-0; I-0**

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

##### Rationale:

- The Committee noted that this measure is used in PQRS and will be available for public reporting on Physician Compare in late 2017.
- No unintended consequences have been identified. Nonetheless, the Committee noted a potential consequence of decreasing bone scan testing rates would be undiagnosed metastatic

|  |
|--|
| <b>0389 Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients</b>   |
| disease; however, this is unlikely based on the evidence for low-risk prostate cancer patients.  |
| <b>5. Related and Competing Measures</b> <ul style="list-style-type: none"> <li>This measure is related to:           <ul style="list-style-type: none"> <li>#0390: Prostate Cancer: Adjuvant Hormonal Therapy for High or Very High Risk Prostate Cancer Patients</li> <li>#1853: Radical Prostatectomy Pathology Reporting</li> </ul> </li> <li>Measures #0390 and #1853 assess different target populations and different aspects of prostate cancer care.</li> </ul> |
| <b>Standing Committee Recommendation for Endorsement: Y-19; N-0</b>  |
| <b>6. Public and Member Comment</b> <ul style="list-style-type: none"> <li></li> </ul>   |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>   |
| <b>8. Board of Directors Vote: Y-X; N-X</b>  |
| <b>9. Appeals</b>  |

## 2963 Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients – Legacy eMeasure

[Submission](#) | [Specifications](#)

**Description:** Percentage of patients, regardless of age, with a diagnosis of prostate cancer at low (or very low) risk of recurrence receiving interstitial prostate brachytherapy, OR external beam radiotherapy to the prostate, OR radical prostatectomy, OR cryotherapy who did not have a bone scan performed at any time since diagnosis of prostate cancer

**Numerator Statement:** Patients who did not have a bone scan performed at any time since diagnosis of prostate cancer

**Denominator Statement:** All patients, regardless of age, with a diagnosis of prostate cancer at low (or very low) risk of recurrence, receiving interstitial prostate brachytherapy, OR external beam radiotherapy to the prostate, OR radical prostatectomy, OR cryotherapy

**Exclusions:** Documentation of medical reason(s) for having a bone scan performed (including documented pain, salvage therapy, other medical reasons)

Documentation of system reason(s) for having a bone scan performed (including bone scan ordered by someone other than reporting physician)

**Adjustment/Stratification:**

**Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Clinician : Team

**Setting of Care:** Ambulatory Care : Ambulatory Surgery Center (ASC), Ambulatory Care : Clinician Office/Clinic, Other

**Type of Measure:** Process

**Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry

**Measure Steward:** PCPI

### STANDING COMMITTEE MEETING [05/18-19/2016]

#### 1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Measure #0389 Evidence Criteria Evaluation Accepted**; 1b. Performance Gap: **Measure #0389 Performance Gap Criteria Evaluation Accepted**

Rationale:

- This “legacy” eMeasure is the eCQM version of the registry measure [#0389](#), currently used in federal programs. The Committee discussed [#0389](#) first, and because the information provided for evidence and opportunity for improvement is identical for the 2 measures, the Committee agreed to assign the ratings for these criteria to #2963.

#### 2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

Initial 2a. Reliability: **M-10; L-6; I-4** 2b. Validity: **M-16; L-2; I-2**

Re-vote on 2a. Reliability: **M-14; L-2; I-1**

Rationale:

- The developer conducted data element validity testing using 34 synthetic patients created in the Bonnie testing system simulating the year 2012. This testing method is appropriate for Legacy eMeasures and satisfies the reliability testing requirement. The Bonnie testing tool was used to test the numerator, denominator, exceptions, measure logic, and value sets to ensure the measure performs as expected. The Bonnie testing results demonstrated 100% coverage and 100% passing rate confirming there was a test case for each pathway of logic and each test case performed as expected.
- The developer provided reliability results from the registry measure ([#0389](#)) and stated that once

## 2963 Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients – Legacy eMeasure

data from the eCQM are available for analysis it is expected that reliability test results will be comparable for the 2 measures. The Committee questioned extrapolating the reliability of the eCQM based on the registry measure without testing results. The Committee questioned if the developer had tested the correlation of the eCQM and registry measure. The developer clarified that although the eCQM is currently used in Meaningful Use (MU), CMS has not released performance data from MU. The Committee noted their concerns with providers' ability to consistently implement the Health Quality Measure Format (HQMF) specifications for the eCQM and the potential impact on the numerator, denominator, and exceptions.

- The Committee acknowledged the importance of eMeasures and the challenges associated with respecifying registry and claims measures and encouraged CMS to release MU performance data.
- The developer conducted face validity testing with a panel of 17 experts representing the PCPI Measures Advisory Committee. Eighty percent of the total respondents (10) either agreed or strongly agreed that the measure can accurately distinguish good and poor quality. The Committee agreed the validity testing results were sufficient.

### 3. Feasibility: H-5; M-13; L-1; I-1

*(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)*

#### Rationale:

- The developer provided information on feasibility testing in the eMeasure Feasibility Score Card for 2 implementation sites and an explanation for scores below 2 on a scale from 1 to 3. Bonnie testing verified that the measure logic is functional, but not all of the required data elements exist as structured data in the unidentified EHRs that were used for testing feasibility. The Committee agreed that the developer provided sufficient information to demonstrate feasibility.

### 4. Usability and Use: H-9; M-11; L-0; I-0

*(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)*

#### Rationale:

- The measure is currently used in Meaningful Use Stage 2 (EHR Incentive Program).
- No unintended consequences have been identified, but similar to #0389, the Committee noted a potential consequence of decreasing bone scan testing rates would be higher rates of undiagnosed metastatic disease. However, this is unlikely based on the evidence for low-risk prostate cancer patients.

### 5. Related and Competing Measures

- This measure is related to:
  - #0390: Prostate Cancer: Adjuvant Hormonal Therapy for High or Very High Risk Prostate Cancer Patients (AMA-PCPI)
  - #1853: Radical Prostatectomy Pathology Reporting (CAP)
- The developer stated that the measure specifications are not completed harmonized; #0390 and #1853 address different target populations and different aspects of prostate cancer care.

### Standing Committee Recommendation for Endorsement: Y-20; N-0

### 6. Public and Member Comment

- After the comment period, the Committee emphasized their concerns with the lack of data from the measure as specified. The developer agreed to provide the Standing Committee with additional data during the measure's scheduled annual review (within one year). The Standing Committee recommended the measure on the condition the measure is reviewed through an ad-hoc review (after scheduled annual review).

|  |
|--|
| <b>2963 Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients – Legacy eMeasure</b> |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>   |
| <b>8. Board of Directors Vote: Y-X; N-X</b>  |
| <b>9. Appeals</b>  |

### 0390 Prostate Cancer: Adjuvant Hormonal Therapy for High or Very High Risk Prostate Cancer Patients

[Submission](#) | [Specifications](#)

**Description:** Percentage of patients, regardless of age, with a diagnosis of prostate cancer at high or very high risk of recurrence receiving external beam radiotherapy to the prostate who were prescribed adjuvant hormonal therapy (GnRH [gonadotropin-releasing hormone] agonist or antagonist)

**Numerator Statement:** Patients who were prescribed adjuvant hormonal therapy (GnRH [gonadotropin-releasing hormone] agonist or antagonist)

**Denominator Statement:** All patients, regardless of age, with a diagnosis of prostate cancer at high or very high risk of recurrence receiving external beam radiotherapy to the prostate

**Exclusions:** AUA methodology uses three categories of reasons for which a patient may be excluded from the denominator of an individual measure. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a medical, patient, or system reason. Examples are provided in the measure exception language of instances that may constitute an exception and are intended to serve as a guide to clinicians. For this measure, exceptions for not prescribing/administering adjuvant hormonal therapy may include medical reason(s) (eg, salvage therapy) or patient reason(s). Although this methodology does not require the external reporting of more detailed exception data, the AUA recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The AUA also advocates the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement. For example, it is possible for implementers to calculate the percentage of patients that physicians have identified as meeting the criteria for exception. Additional details by data source are as follows:

Documentation of medical reason(s) for not prescribing/administering adjuvant hormonal therapy (eg, salvage therapy)

Documentation of patient reason(s) for not prescribing/administering adjuvant hormonal therapy

**Adjustment/Stratification:**

**Level of Analysis:** Clinician : Group/Practice, Clinician : Individual, Clinician : Team

**Setting of Care:** Ambulatory Care : Ambulatory Surgery Center (ASC), Ambulatory Care : Clinician Office/Clinic, Other

**Type of Measure:** Process

**Data Source:** Electronic Clinical Data, Electronic Clinical Data : Registry

**Measure Steward:** American Urological Association

#### STANDING COMMITTEE MEETING [05/18-19/2016]

##### 1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Previous Evidence Evaluation Accepted**; 1b. Performance Gap: **H-1; M-13; L-5; I-1**

##### Rationale:

- For the 2012 endorsement evaluation, the developer provided clinical practice guidelines from the American Urological Association (AUA) and the National Comprehensive Cancer Network (NCCN) stating physicians should consider the use of external beam radiotherapy and concurrent use of hormonal therapy in high-risk prostate cancer patients to prolong survival.
- For the current evaluation, the developer provided updates to the guidelines, and included a Cochrane Review of the body of evidence. There were no changes to the recommendations since the previous submission. The Committee agreed that the updated evidence supports the measure focus and has a stronger level of evidence. The Committee accepted the prior evaluation of this criterion without further discussion.

### 0390 Prostate Cancer: Adjuvant Hormonal Therapy for High or Very High Risk Prostate Cancer Patients

- The developer provided average performance rates from the Physician Quality Reporting System (PQRS) Registry from 2010 – 2013. The average performance rate was 79.6% in 2010, 93.5% in 2011, 91.1% in 2012, and 95.4% in 2013. The mean performance rate in 2014 was 93.82%, the minimum was 16.67%, and the maximum was 100.0%. The developer did not provide data on disparities from the measure as specified but provided evidence from the literature that demonstrated higher incidence rates of prostate cancer in African-American men compared to white men. The literature also showed that African-American men are more likely to receive non-surgical treatment than white men and white men were less likely than African-American men to receive radiation therapy and hormonal therapy.
- The developer also cited an analysis of data from the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) registry that found that the utilization of adjuvant hormonal therapy and external beam radiotherapy for high-risk patients has increased to 80.0% throughout the past two decades, yet utilization rates have plateaued since 2000.
- The Committee agreed that performance has improved over time but there is still an opportunity for improvement.

#### **2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria**

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Previous Reliability Evaluation Accepted** 2b. Validity: **M-14; L-5; I-1**

##### Rationale:

- For the 2012 endorsement evaluation, inter-rater reliability was conducted on 91 patient records from 2010; chart and data auditing occurred in 2011. The percent agreement for the numerator, denominator, and exclusions/exceptions was 100.0% and 98.9% for overall reliability.
- For the current evaluation, the developer provided updated reliability testing of the measure score using a beta-binomial model to assess the signal-to-noise ratio. Reliability at the minimum level of quality reporting events (10) was 0.73 and 0.85 at the average number of quality events (21.5). A reliability of 0.70 is generally considered a minimum threshold for reliability.
- The Committee agreed that the updated reliability testing results were satisfactory and accepted the prior evaluation of this criterion without further discussion.
- For the 2012 endorsement evaluation, face validity of the measure score as an indicator of quality was systematically assessed by an expert panel. The expert panel agreed that the scores obtained from the measure as specified provide an accurate reflection of quality and can be used to distinguish good and poor quality.
- For the current evaluation, the developer conducted additional face validity testing with a panel of 21 experts representing the AUA Committee on Quality Improvement and Patient Safety. One hundred percent of the total respondents (15) either agreed or strongly agreed that this measure can accurately distinguish good and poor quality.
- The Committee discussed the seemingly excessive exclusion/exception rate of 204 exclusions/exceptions per 20 providers with an overall rate of 32.2%. One of the Committee members noted that although some patients should not receive adjuvant hormonal therapy, the exclusion/exception rate appeared relatively high. The Committee also questioned the usefulness of the measure since one-third of the patients were excluded.
- Despite the potential threat to validity due to the high rate of exclusion/exception rates, the Committee agreed that without specific information about the exclusions/exceptions, it was difficult to understand how validity of the measure overall was impacted; therefore, the updated validity testing results were sufficient and met the validity criterion.

#### **3. Feasibility: H-9; M-10; L-0; I-1**

|   |
|---|
| <b>0390 Prostate Cancer: Adjuvant Hormonal Therapy for High or Very High Risk Prostate Cancer Patients</b>  |
| <i>(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)</i>  |
| <p><u>Rationale:</u></p> <ul style="list-style-type: none"> <li>The Committee did not note any concerns regarding feasibility, acknowledging that the data elements used to construct this measure are based on clinical registry data and available in electronic sources.</li> </ul>  |
| <p><b>4. Usability and Use: H-8; M-11; L-1; I-0</b></p> <p><i>(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)</i></p> <p><u>Rationale:</u></p> <ul style="list-style-type: none"> <li>The Committee noted that this measure is used in the Physician Quality Reporting System (PQRS) and is also used in the AUA Quality (AQUA) Registry.</li> </ul>  |
| <p><b>5. Related and Competing Measures</b></p> <ul style="list-style-type: none"> <li>This measure is related to: <ul style="list-style-type: none"> <li>0220: Adjuvant hormonal therapy</li> <li>0389: Prostate Cancer: Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients</li> <li>1853: Radical Prostatectomy Pathology Reporting</li> </ul> </li> <li>According to the developer the measures specifications are not completely harmonized. Measure #0220 focuses on adjuvant hormonal therapy for breast cancer patients. Measures #0389 and #1853 have different target populations and address different aspects of prostate cancer care.</li> </ul> |
| <b>Standing Committee Recommendation for Endorsement: Y-19; N-1</b>   |
| <p><b>6. Public and Member Comment</b></p> <ul style="list-style-type: none"> <li></li> </ul>   |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>  |
| <b>8. Board of Directors Vote: Y-X; N-X</b>   |
| <b>9. Appeals</b>   |

## 0508 Diagnostic Imaging: Inappropriate Use of “Probably Benign” Assessment Category in Screening Mammograms

[Submission](#) | [Specifications](#)

**Description:** Percentage of final reports for screening mammograms that are classified as “probably benign”

**Numerator Statement:** Final reports classified as “probably benign”

**Denominator Statement:** All final reports for screening mammograms

**Exclusions:** No Denominator Exclusions or Denominator Exceptions

**Adjustment/Stratification:**

**Level of Analysis:** Clinician : Individual

**Setting of Care:** Ambulatory Care : Clinician Office/Clinic, Imaging Facility

**Type of Measure:** Process

**Data Source:** Administrative claims, Electronic Clinical Data : Registry

**Measure Steward:** American College of Radiology (ACR)

### STANDING COMMITTEE MEETING [05/18-19/2016]

#### 1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **H-0; M-7; L-0; I-13**; 1b. Performance Gap: **H-3; M-14; L-4; I-0**; Evidence Exception: **Y-20; N-1**

#### Rationale:

- For the 2008 endorsement evaluation,<sup>5</sup> the developer provided a guideline recommendation from the American College of Radiology (ACR) Breast Imaging Reporting Data System (BI-RADS®) Atlas, 2003 that stated: Do not use “probably benign” (Category 3) in interpreting screening examinations (level of evidence is not graded).
- For the current evaluation, the developer provided several sources of evidence that did not support the measure focus including an updated recommendation from the ACR BI-RADS® 5<sup>th</sup> edition, 2012 that recommends overall final assessment of findings should be based on all imaging studies performed up to that day. In addition, they must be classified according to the FDA-approved final assessment categories and should follow the define categories (level of evidence is not graded). The developer also provided a recommendation from the U.S. Preventive Services Task Force (USPSTF) that included biennial screening mammography for women within different age groups and risks. The developer, did however, provide 8 studies from the literature addressing the “probably benign” category.
- Based on NQF criteria, the evidence was insufficient due to lack of empirical evidence provided to support this process of care: “probably benign” should not be used as a category for indeterminate findings. The Committee agreed that the evidence was insufficient but that it is beneficial to hold providers accountable for performance in the absence of empirical evidence of benefits to patients.
- The developer provided physician performance rates from the CMS Physician Quality Reporting System (PQRS) from 2012 – 2014. The performance rate in 2012 was 2.09%, 5.48% in 2013, and 0.49% in 2014. The goal of this measure is a zero-reporting rate. The developer did not provide data on disparities from the measure as specified.
- The Committee agreed that based on the performance data provided by the developer, providers were still using the “probably benign” assessment category 0.49% of the time in 2014, therefore, an opportunity for improvement still exists.

#### 2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Previous Reliability Evaluation Accepted** 2b. Validity: **Previous Validity Evaluation**

## 0508 Diagnostic Imaging: Inappropriate Use of “Probably Benign” Assessment Category in Screening Mammograms

### Accepted;

#### Rationale:

- For the 2012 endorsement evaluation,<sup>6</sup> inter-rater reliability was conducted on 114 patient records from 3 radiology practices from 2010. The percent agreement for the numerator, denominator, and overall reliability was 100.0%.
- For the current evaluation, the developer provided updated reliability testing of the measure score using a beta-binomial model to assess the signal-to-noise ratio. The mean reliability was 0.99. A reliability of 0.70 is generally considered a minimum threshold for reliability.
- The Committee agreed that the updated reliability testing results were satisfactory and accepted the prior evaluation of this criterion without further discussion.
- For the 2012 endorsement evaluation,<sup>7</sup> face validity of the measure score as an indicator of quality was systematically assessed by an expert panel. The expert panel agreed that the scores obtained from the measure as specified provide an accurate reflection of quality and can be used to distinguish good and poor quality.
- For the current evaluation, the developer conducted additional face validity testing with a panel of 20 experts representing the ACR Commission on Breast Imaging and the National Mammography Database. Eleven respondents either agreed or strongly agreed that physicians who perform well on this measure demonstrate a higher level of quality than physicians who do not perform well on this measure.
- The Committee agreed that the updated validity testing results were satisfactory and accepted the prior evaluation of this criterion without further discussion.

### 3. Feasibility: H-19; M-1; L-1; I-0

*(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)*

#### Rationale:

- The Committee did not note any concerns regarding feasibility, acknowledging that the data elements used to construct this measure are based on clinical registry data and available in electronic sources.

### 4. Usability and Use: H-17; M-3; L-1; I-0

*(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)*

#### Rationale:

- The Committee noted that this measure is used in the Physician Quality Reporting System (PQRS) and Value Based Payment Modifier.

### 5. Related and Competing Measures

- No related or competing measures noted.

### Standing Committee Recommendation for Endorsement: Y-20; N-1

### 6. Public and Member Comment

- 

### 7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X

### 8. Board of Directors Vote: Y-X; N-X

### 9. Appeals

## 0509 Diagnostic Imaging: Reminder System for Screening Mammograms

[Submission](#) | [Specifications](#)

**Description:** Percentage of patients undergoing a screening mammogram whose information is entered into a reminder system with a target due date for the next mammogram

**Numerator Statement:** Patients whose information is entered into a reminder system with a target due date for the next mammogram

**Denominator Statement:** All patients undergoing a screening mammogram

**Exclusions:** Documentation of medical reason(s) for not entering patient information into a reminder system [(eg, further screening mammograms are not indicated, such as patients with a limited life expectancy, other medical reason(s))]

**Adjustment/Stratification:**

**Level of Analysis:** Clinician : Individual

**Setting of Care:** Hospital/Acute Care Facility, Imaging Facility

**Type of Measure:** Process

**Data Source:** Administrative claims, Electronic Clinical Data : Registry

**Measure Steward:** American College of Radiology

### STANDING COMMITTEE MEETING [05/18-19/2016]

#### 1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Previous Evidence Evaluation Accepted**; 1b. Performance Gap: **H-5; M-14; L-1; I-1**

#### Rationale:

- For the 2008 endorsement evaluation,<sup>8</sup> the developer provided a guideline recommendation from the American College of Radiology (ACR) Breast Imaging Reporting Data System (BI-RADS®) Atlas, 2003 that stated: Do not use 'probably benign' (Category 3) in interpreting screening examinations (level of evidence is not graded).
- For the current evaluation, the developer provided a recommendation from the Community Preventive Services Task Force that recommends the use of client reminders to increase screening for breast and cervical cancers on the basis of strong evidence of effectiveness.
- The Committee agreed that the updated evidence provided was stronger than the previous evidence and accepted the prior evaluation of this criterion without further discussion.
- The developer provided physician performance rates from the CMS Physician Quality Reporting System (PQRS) from 2012 – 2014. The performance rate in 2012 was 79.4%, 86.0% in 2013, and 87.6% in 2014. The developer did not provide data on disparities from the measure as specified but cited a 2010 National Health Interview Survey that demonstrated Asian race, low education status, recent immigrant status, and no regular source of medical care or no medical insurance were factors found to reduce the likelihood for a woman to receive a mammogram.
- The Committee agreed that based on the performance data provided by the developer, an opportunity for improvement still exists.

#### 2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Previous Reliability Evaluation Accepted** Initial 2b. Validity: **M-9; L-7; I-5**

Re-vote on 2b. Validity: **M-13; L-7**

#### Rationale:

- For the 2012 endorsement evaluation,<sup>9</sup> inter-rater reliability was conducted on 114 patient records from 3 radiology practices from 2010. The percent agreement for the numerator, denominator, and overall reliability was 100.0%.
- For the current evaluation, the developer provided updated reliability testing of the measure

### 0509 Diagnostic Imaging: Reminder System for Screening Mammograms

score using a beta-binomial model to assess the signal-to-noise ratio. The mean reliability was 0.88. A reliability of 0.70 is generally considered a minimum threshold for reliability.

- The Committee agreed that the updated reliability testing results were satisfactory and accepted the prior evaluation of this criterion without further discussion.
- For the 2012 endorsement evaluation,<sup>10</sup> face validity of the measure score as an indicator of quality was systematically assessed by an expert panel. The expert panel agreed that the scores obtained from the measure as specified provide an accurate reflection of quality and can be used to distinguish good and poor quality.
- For the current evaluation, the developer conducted additional face validity testing with a panel of 20 experts representing the ACR Commission on Breast Imaging and the National Mammography Database. Ten respondents generally agreed that physicians who perform well on this measure demonstrate a higher level of quality than physicians who do not perform well on this measure.
- The exclusion, ‘medical reason documentation’ was added in 2014; however, the developer did not conduct an analysis to determine the impact of this exclusion on the validity of the measure. The developer stated that the exclusion allows physicians to report on the measure if a patient’s information was not entered into a reminder system because it was determined that they did not need to return for a screening mammogram due to decreased life expectancy, history of a mastectomy, or some other medical reason. The developer explained that the exclusion should not be a threat to validity because it was only used 3 times during 2014. Committee members then questioned why the exclusions were so low considering that the developer was reporting Medicare data from PQRS and expected the number of exclusions to be higher in the Medicare population. This raised concerns about the exclusion not being used properly by physicians and the need for the exclusion. During the post-comment call, the developer stated that they would analyze the 2015 PQRS data and consider removing the exclusion. The developer will provide the additional data analysis during the measure’s annual review (within one year) for the Committee’s review. The Committee re-voted on the validity of the measure and recommended it for endorsement.

#### 3. Feasibility: H-0; M-19; L-0; I-2

*(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)*

##### Rationale:

- This measure is based on clinical registry data and all data elements are available in electronic sources.
- While the Committee did recommend having an age range for women in the denominator, they agreed the measure was feasible.

#### 4. Usability and Use: H-1; M-18; L-2; I-0

*(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)*

##### Rationale:

- The Committee noted that this measure is used in PQRS and is also used for quality improvement with benchmarking in the ACR NRDR Qualified Clinical Data Registry.

#### 5. Related and Competing Measures

- This measure is related to:
  - #2372 : Breast Cancer Screening (NQCA)
- The developer stated that the measures have the same measure focus and target population. According to the developer, the measure specifications are completely harmonized.

**Standing Committee Recommendation for Endorsement: Y-18; N-3**

|  |
|--|
| <b>0509 Diagnostic Imaging: Reminder System for Screening Mammograms</b>                                       |
| <b>6. Public and Member Comment</b> <ul style="list-style-type: none"><li>No comments were received.</li></ul> |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>   |
| <b>8. Board of Directors Vote: Y-X; N-X</b>  |
| <b>9. Appeals</b>  |

**0559 Combination chemotherapy is recommended or administered within 4 months (120 days) of diagnosis for women under 70 with AJCC T1cN0M0, or Stage IB - III hormone receptor negative breast cancer.**

[Submission](#) | [Specifications](#)

**Description:** Percentage of female patients, age >18 at diagnosis, who have their first diagnosis of breast cancer (epithelial malignancy), at AJCC stage T1cN0M0 (tumor greater than 1 cm), or Stage IB -III, whose primary tumor is progesterone and estrogen receptor negative recommended for multiagent chemotherapy (recommended or administered) within 4 months (120 days) of diagnosis.

**Numerator Statement:** Combination chemotherapy is administered within 4 months (120 days) of the date of diagnosis or it is recommended and not received.

**Denominator Statement:** Women under the age of 70 with AJCC T1cN0M0, or Stage IB-III hormone receptor negative breast cancer:

- Women
- Age 18-69 at time of diagnosis
- Known or assumed first or only cancer diagnosis
- Primary tumors of the breast
- Epithelial invasive malignanc

**Exclusions:** Exclude, if any of the following characteristics are identified:

Men;

Age <18 and >=70;

not a first or only cancer diagnosis;

non-epithelial and non-invasive tumors;

phyllodes tumor histology;

rare histology not supported by clinical trials: 8940 - Mixed tumor, malignant, NOS, 8950 - Mullerian mixed tumor, 8980 – Carcinosarcoma, 8981 - Carcinosarcoma, embryonal

Tumor size <=1cm and AJCC pN=0;

ERA positive;

PRA positive;

Evidence of in situ or metastatic disease;

Not treated surgically;

Died within 4 months (120 days) of diagnosis;

Participation in a clinical trial which directly impacts the delivery of the standard of care

**Adjustment/Stratification:**

**Level of Analysis:** Facility

**Setting of Care:** Hospital/Acute Care Facility

**Type of Measure:** Process

**Data Source:** Paper Medical Records, Electronic Clinical Data : Registry

**Measure Steward:** Commission on Cancer, American College of Surgeons

**STANDING COMMITTEE MEETING [05/18-19/2016]**

**1. Importance to Measure and Report: The measure meets the Importance criteria**

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Previous Evidence Evaluation Accepted**; 1b. Performance Gap: **H-7; M-12; L-1; I-0**

Rationale:

- For the 2012 endorsement evaluation, the developer provided a clinical practice guideline from the National Comprehensive Cancer Network (NCCN) as evidence to support the administration

**0559 Combination chemotherapy is recommended or administered within 4 months (120 days) of diagnosis for women under 70 with AJCC T1cN0M0, or Stage IB - III hormone receptor negative breast cancer.**

of combination chemotherapy to breast cancer patients whose primary tumor is progesterone and estrogen receptor negative. The developer also included a systematic review of the body of evidence with multiple randomized clinical trials demonstrating approximately 33.0% reduction in risk of distant cancer recurrence and death.

- The Committee agreed that the evidence basis for the measure has not changed and there was no need to repeat the discussion and vote on evidence.
- For the current evaluation, the developer provided national trend data from the National Cancer Data Base (NCDB) from 2008 and 2013. The mean performance rate for 2008 was 85.1% and 89.4% for 2013.
- The Committee agreed that based on the performance and disparities data provided by the developer, a gap in care continues to exist in the administration of combination chemotherapy for breast cancer patients.
- The Committee suggested monitoring the impact of emerging breast cancer data and new genomic assays that may potentially exclude patients with hormone receptor negative tumors from receiving chemotherapy.

**2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria**

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

Initial 2a. Reliability: **M-9; L-9; I-2** 2b. Validity: **M-9; L-9; I-2**

Re-vote on 2a. Reliability: **H-2; M-12** 2b. Validity: **H-2; M-12**

**Rationale:**

- For the 2012 endorsement evaluation, the developer provided mean performance rates that included 1,400 CoC-accredited cancer programs and approximately 14,000 cases from 2007 (86.3) and 2008 (84.9).
- For the current evaluation, the developer provided updated performance rates from 2013 showing the hospital-level performance rates from 0% to 100%. NQF reliability testing requirements include statistical analysis of the computed measure score or the individual patient-level data for the measured entities to determine the proportion of variation due to true differences vs. noise or random variation. Overall performance rates do not meet the reliability criterion, which was provided by the developer. Data element validity testing was performed and counted for data element reliability.
- For the 2012 endorsement evaluation, validity was assessed by randomly selecting charts and reviewing them by site surveyors to determine completeness and validity of data reported to registry. The measure numerator and denominator were viewed by the clinical constituency within these cancer programs as valid and an appropriate reflection of the standard of care described in NCCN clinical guidelines.
- For the current evaluation, the developer provided additional details on data element validity testing conducted in 2009 and 2010 by comparing registry data to data that were re-abstracted from the medical records by CoC site surveyors, which was considered the gold standard. The developer provided percentage agreement results for 2 of the data elements included in the numerator, timing for chemotherapy (81.1, 75.7) and chemotherapy which was recommended but not administered (88.1, 89.5). Validity testing of all the critical data elements (including kappa scores, sensitivity or specificity statistics) was not provided.
- Since the testing provided by the developer for this measure had the same issues as [#0220](#), the Committee considered the same concerns they had for the testing of that measure and agreed to carry forward the votes from the reliability and validity criteria from [#0220](#) and recommended the measure for endorsement
- The Committee encouraged the developer to provide updated reliability and validity testing at

**0559 Combination chemotherapy is recommended or administered within 4 months (120 days) of diagnosis for women under 70 with AJCC T1cN0M0, or Stage IB - III hormone receptor negative breast cancer.**

the next maintenance review of the measure. The developer confirmed that they are planning to update their validity and reliability testing for the 5 measures submitted in this project (#0219, #0220, #0223, #0225, and #0559).

- The Committee agreed there may be multiple providers and procedures (genetic testing, surgery, etc.) from the time of diagnosis to the start of chemotherapy that may extend beyond the 120 day timeframe required by the measure, but facilities should aim to prevent delays in initiating treatment and improving patient outcomes.

**3. Feasibility: H-4; M-14; L-2; I-0**

*(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)*

Rationale:

- Although the data are readily available through medical records, the Committee recognized the data collection burden for manual chart abstraction that could result in various interpretations.
- The Committee agreed this measure meets the feasibility criterion.

**4. Usability and Use: H-9; M-9; L-2; I-0**

*(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)*

Rationale:

- The measure is publicly reported through the Pennsylvania Health Care Quality Alliance, the Commission on Cancer, various compliance benchmarking programs through the National Cancer Data Base, and Quality Oncology Practice Initiative programs.
- The developer provided improvement results showing increases in the overall facility level compliance rates and across all census regions.
- The Committee agreed this measure meets the usability and use criterion.

**5. Related and Competing Measures**

- No related or competing measures noted.

**Standing Committee Recommendation for Endorsement: Y-19; N-1**

**6. Public and Member Comment**

- One commenter stated that it would be beneficial to have the measure stipulate administered vs. prescribed and to address who might not receive the treatment via the exclusions.
- Developer response: The language of “recommended or administered” in these measures was specifically selected after discussion with clinicians and users and is based directly on the FORDs data item definitions used to calculate these measures. We agree with that when assessing overall quality, cancer programs should review patients in which treatment is administered and those in which treatment is recommended but not administered. Therefore, in the our reporting systems where compliance with these measures is assessed, cancer programs are able to view cases stratified by if; a) treatment is administered, b) treatment is recommended but not administered and c) the case is non-compliant with the measure. This allows programs to assess patients which cases are compliant with the measure but for which adjuvant therapy was not administered during internal quality improvement efforts.
- During the Comment period, the developer submitted additional performance data from the Rapid Quality Reporting System (RQRS). The developer stated that the RQRS performance rates were similar to the performance rates from the NCDB.
- The Committee considered the additional performance data from the Rapid Quality Reporting System (RQRS) and agreed this was an important indicator for cancer care. On re-vote, the

|   |
|---|
| 0559 Combination chemotherapy is recommended or administered within 4 months (120 days) of diagnosis for women under 70 with AJCC T1cN0M0, or Stage IB - III hormone receptor negative breast cancer. |
| Committee recommended the measure for endorsement.  |
| 7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X   |
| 8. Board of Directors Vote: Y-X; N-X  |
| 9. Appeals  |

## 2930 Febrile Neutropenia Risk Assessment Prior to Chemotherapy

[Submission](#) | [Specifications](#)

**Description:** Percentage of patients with a solid malignant tumor or lymphoma who had a febrile neutropenia (FN) risk assessment completed and documented in the medical record prior to the first cycle of intravenous chemotherapy

**Numerator Statement:** Number of patients who had an FN risk assessment documented in the medical record prior to the first cycle of intravenous chemotherapy.

**Denominator Statement:** Number of patients 18 years of age or older with a solid malignant tumor or lymphoma receiving the first cycle of intravenous chemotherapy.

**Exclusions:** There are no denominator exclusions.

**Adjustment/Stratification:**

**Level of Analysis:** Clinician : Group/Practice

**Setting of Care:** Ambulatory Care : Clinician Office/Clinic, Other

**Type of Measure:** Process

**Data Source:** Electronic Clinical Data : Electronic Health Record, Paper Medical Records

**Measure Steward:** RAND Corporation

### STANDING COMMITTEE MEETING [05/18-19/2016]

#### 1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **H-4; M-15; L-0; I-0**; 1b. Performance Gap: **H-3; M-12; L-2; I-2**

#### Rationale:

- The developer provided a clinical practice guideline from the 2015 American Society of Clinical Oncology (ASCO) Recommendations for the Use of WBC Growth Factors and the 2015 NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) to support the assessment of febrile neutropenia (FN) risk and administration of appropriate colony-stimulating factor (CSF) prophylaxis prior to chemotherapy. The developer provided additional studies evaluating the effectiveness of FN risk assessment tools. The Committee noted that the developer presented strong evidence supporting the administration of CSF prophylaxis prior to chemotherapy. However, the focus of the measure is documentation of a FN risk assessment prior to chemotherapy. The developer clarified that there is no evidence supporting 1 FN risk assessment tool over another at this time. The Committee agreed the evidence the developer provided to support the use of a FN assessment tool demonstrated a decrease in the incidence of febrile neutropenia and related complications.
- The developer provided performance rates from April 2011-February 2016 that included 192 patient records from 5 community oncology clinics. The mean performance rate was 12.0%, the median was 16.0%, and the maximum was 27.0%. The performance rates were stratified by age, race/ethnicity, and gender. The developer provided data from the literature that showed disparities on the use of prophylactic CSF based on gender, race, geographic location, and lower socioeconomic status. The developer stated that there is limited published data on the frequency of risk assessment for FN but cited a study (Miller, 2010) conducted at 4 offices of a community oncology practice to assess the effect of a computer-based risk assessment tool (CBRAT) for FN. Before implementation of the CBRAT, 13 of 101 (13.0%) patients had documented risk assessments for FN. After implementation of CBRAT, documented risk assessments increased to 100.0%.
- The Committee noted that appropriately administering prophylactic CSF and preventing FN in high-risk cancer patients is important, but based on the limited data the developer provided, the Committee questioned whether a gap in care/quality problem exists related to documentation of a FN assessment. The Committee suggested that the low performance rates presented by the

### 2930 Febrile Neutropenia Risk Assessment Prior to Chemotherapy

developer may be due to the adoption of computerized physician order entry (CPOE) and standard order sets that include supportive care treatments appropriate for the regimen, including pre-medications, hydration, CSF, and hypersensitivity medications. Providers using standardized orders sets are not likely to include additional documentation explicitly stating the FN risk or a note in the chart that reflects the rationale for either administering or not administering CSF based on patient and regimen risk factors as required by the measure.

- The Committee agreed that it is important to assess patients for FN risk and administer CSF appropriately, however, they encouraged the developer to expand the measure so that evidence-based standing orders meet the intent of the measure.

#### 2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **M-14; L-4; I-1** 2b. Validity: **H-17; M-0; L-1**

##### Rationale:

- The Committee agreed the data elements are clearly defined, but somewhat complex and may be difficult to calculate consistently.
- Inter-rater reliability testing was assessed using 2 abstractors who were instructed to abstract the same randomly selected 50 medical records from 5 community oncology clinics, 10 records per clinic, for a 25 percent inter-rater reliability (IRR) sample. The kappa statistic and percent agreement between the abstractors was calculated based on whether documentation of a febrile neutropenia risk assessment was in the medical record. The developer provided kappa statistics and percent agreement results for 1 data element included in the numerator (documentation of a febrile neutropenia risk assessment in the medical record). Kappa estimates ranged from 0.783 to 1.0 for the 5 clinics; percent agreement ranged from 90-100%. NQF guidance states that testing should be done for all critical data elements. The clinics determined which patients met the denominator inclusion criteria (age at least 18 years, solid tumor or lymphoma, initiating chemotherapy, and not participating in a clinical trial). The developers excluded additional patients due to incomplete records, malignancy other than solid tumor or lymphoma, or concurrent radiation. The Committee commented that the sample used for reliability testing was relatively small, yet the reliability score was acceptable and met the reliability criterion.
- The Committee encouraged the developer to conduct a statistical analysis, in the future, of the computed measure score to assess the proportion of variability due to real differences among the measured entities.
- The developer assessed face validity of the measure score using a panel of 10 experts in clinical oncology. Eighty percent (8/10) of the respondents either agreed or strongly agreed that performance scores resulting from the measure as defined can be used to distinguish good and poor quality. One of the Committee members commented that they would like to see data showing that groups with high scores on the measure have less FN. Another Committee member suggested that missing data may be a threat to validity, although the developer stated that missing data was not identified during the medical record abstraction. The Committee concluded that the validity criterion was met.

#### 3. Feasibility: **H-1; M-14; L-3; I-0**

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)

##### Rationale:

- Some of the data elements are easily found in electronic sources but information about FN risk assessment may not be generated during routine care delivery and require manual chart abstraction. The Committee suggested incorporating the FN risk assessment into CPOE and standard orders to increase the feasibility of the measure in the future.

|   |
|---|
| <b>2930 Febrile Neutropenia Risk Assessment Prior to Chemotherapy</b>   |
| <p><b>4. Usability and Use: H-1; M-16; L-1; I-0</b></p> <p><i>(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)</i></p> <p><u>Rationale:</u></p> <ul style="list-style-type: none"> <li>The developer stated that because the measure is being submitted to NQF for initial endorsement, they do not yet have plans to submit it for use in a specific federal, state or local program. However, the measure would be appropriate for use in a CMS reporting program for outpatient care provided to oncology patients.</li> <li>The Committee emphasized that a febrile neutropenia outcome measure would further the goal of high-quality, efficient healthcare rather than this process measure. The Committee requested that, if endorsed, the developer provide data on the performance of the measure and include patients who were administered CSF prophylaxis and patients with febrile neutropenia to understand the impact of the measure. Another Committee member questioned the impact this measure will have on the appropriate use of CSF but acknowledged that additional data will be useful to improve quality.</li> </ul> |
| <p><b>5. Related and Competing Measures</b></p> <ul style="list-style-type: none"> <li>No related or competing measures noted.</li> </ul>   |
| <b>Standing Committee Recommendation for Endorsement: Y-16; N-2</b>   |
| <p><b>6. Public and Member Comment</b></p> <ul style="list-style-type: none"> <li>One commenter noted that an outcome measure will assist in determining more than appropriate use of colony-stimulating factor (CSF), specifically resource utilization related to urgent care due to febrile neutropenia (FN). The commenter also noted the challenges of documenting FN risk assessment in electronic health records (EHR).</li> <li>Developer response: We agree that measuring febrile neutropenia (FN) outcomes is important, but view an outcome measure as a complement to our proposed measure rather than a substitute.</li> </ul>  |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>  |
| <b>8. Board of Directors Vote: Y-X; N-X</b>   |
| <b>9. Appeals</b>   |

## Measures Recommended for Inactive Endorsement with Reserve Status

| 1878 HER2 testing for overexpression or gene amplification in patients with breast cancer  |
|--|
| <a href="#">Submission</a>   <a href="#">Specifications</a>  |
| <p><b>Description:</b> Proportion of female patients (aged 18 years and older) with breast cancer who receive human epidermal growth factor receptor 2 (HER2) testing for overexpression or gene amplification</p> <p><b>Numerator Statement:</b> HER2 testing performed</p> <p><b>Denominator Statement:</b> Adult women with breast cancer</p> <p><b>Exclusions:</b> None</p> <p><b>Adjustment/Stratification:</b></p> <p><b>Level of Analysis:</b> Clinician : Group/Practice</p> <p><b>Setting of Care:</b> Ambulatory Care : Clinician Office/Clinic</p> <p><b>Type of Measure:</b> Process</p> <p><b>Data Source:</b> Electronic Clinical Data : Registry</p> <p><b>Measure Steward:</b> American Society of Clinical Oncology</p>   |
| <p><b>STANDING COMMITTEE MEETING [05/18-19/2016]</b></p> <p><b>1. Importance to Measure and Report: <u>The measure did not meet the Importance criteria</u></b><br/>(1a. Evidence, 1b. Performance Gap)</p> <p>1a. Evidence: <b>Previous Evidence Evaluation Accepted</b>; 1b. Performance Gap: <b>H-1; M-3; L-15; I-1</b></p> <p><u>Rationale:</u></p> <ul style="list-style-type: none"> <li>For the 2012 endorsement evaluation, the developer provided a clinical practice guideline from the American Society of Clinical Oncology (ASCO) and the College of American Pathologists (CAP) recommending that human epidermal growth factor receptor 2 (HER2) status should be determined for all invasive breast cancer.</li> <li>For the current evaluation, the developer provided updated citations to the clinical practice guideline but the recommendations did not change.</li> <li>The Committee agreed that the evidence basis for the measure has not changed and there was no need to repeat the discussion and vote on evidence.</li> <li>The developer provided performance rates from the ASCO Quality Oncology Practice Initiative (QOPI®) Registry from 2013 – 2015. The mean performance rate in 2013 was 98.53%, 98.77% in 2014, and 98.63% in 2015. The developer provided disparities data aggregated by race and/or ethnic groups. The developer also noted that studies show that tumors of older female patients (15.7 %) and Hispanics (20.7 %) as well as other race/ethnicities (18.8 %) are less likely to be tested for HER2.</li> <li>The Committee discussed the high performance rates of this measure, noting that there is no longer a gap in performance among the practices being measured. There was discussion about participants in the QOPI Registry being self-selected and voluntarily reporting on this measure and the possibility for practices outside of the registry having lower performance rates. Another Committee member cited Surveillance, Epidemiology, and End Results (SEER) data, which is more nationally representative, from 2007 demonstrating 96.5% of eligible patients had HER2 testing performed.</li> <li>Ultimately, the measure did not pass performance gap. However, despite the high rate of performance there was evidence that disparities exist; therefore, the Committee voted to continue reviewing the measure against the rest of the criteria with the possibility of recommending the measure for inactive endorsement with reserve status.</li> </ul> |
| <p><b>2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u></b><br/>(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)</p> <p>2a. Reliability: <b>Previous Reliability Evaluation Accepted</b> 2b. Validity: <b>Previous Validity Evaluation</b></p>   |

**1878 HER2 testing for overexpression or gene amplification in patients with breast cancer****Accepted;****Rationale:**

- For the 2012 endorsement evaluation, data element validity testing was performed and counted for data element reliability. The dataset included 264 patient records from 44 QOPI practices submitted in spring 2007. Trained, independent nurse abstractors served as the “gold standard” against which practice abstractions were compared for accuracy. Kappa statistics were used to analyze the validity of the audited patient records compared to the submitted patient records. By convention, a kappa > 0.70 is considered acceptable. The developer provided a kappa score of 0.85 and an overall percent agreement of 98.0%. While this kappa score is above what is considered acceptable, the developer did not state which of the data elements this kappa score represented; no additional results were provided. NQF guidance states that testing should be done for all critical data elements. The developer responded that they were unable to find the additional data from the testing previously conducted but based on the kappa score and overall agreement rate they did not have any concerns with the performance of the measure in the registry.
- Although the developer did not provide updated reliability and validity testing, the Committee agreed that the measure specifications were consistently implemented within the registry and accepted the previous reliability and validity evaluation.

**3. Feasibility: H-3; M-15; L-1; I-1**

*(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)*

**Rationale:**

- This measure is based on clinical registry data and all data elements are available in electronic sources. The Committee agreed the measure is feasible.
- The Committee noted that eventual use of this measure through EHRs would lessen the data collection burden.

**4. Usability and Use: H-4; M-13; L-3; I-0**

*(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)*

**Rationale:**

- The developer shared that this measure was recently selected for inclusion in a Medical Oncology Core Measure Set supported by AHIP and CMS. The measure was also recently approved for use in the Medicare Access & CHIP Reauthorization Act’s (MACRA) Merit-Based Incentive Payment System (MIPS).
- The Committee noted that this measure is used in Quality Oncology Practice Initiative (QOPI®), the QOPI® Certification Program, and the PQRS Qualified Clinical Data Registry.

**5. Related and Competing Measures**

- This measure is related to:
  - #1855: Quantitative HER2 Evaluation by IHC uses the System Recommended by the ASCO/CAP Guidelines (CAP)
- Measure #1855 and #1878 address 2 complimentary components and are related to appropriate identification and treatment of breast cancer patients. Measure #1855 and #1878 differ by data source. Measure #1878 is suited for registry data. Measure #1855 is suited for administrative claims and paper medical records data sources. The developer indicates the measures have been harmonized.

**Standing Committee Recommendation for Inactive Endorsement with Reserve Status : Y-19; N-1**

**6. Public and Member Comment**

|  |
|--|
| <b>1878 HER2 testing for overexpression or gene amplification in patients with breast cancer</b> |
| •  |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>                           |
| <b>8. Board of Directors Vote: Y-X; N-X</b>  |
| <b>9. Appeals</b>  |

## 1857 HER2 negative or undocumented breast cancer patients spared treatment with HER2-targeted therapies

[Submission](#) | [Specifications](#)

**Description:** Proportion of female patients (aged 18 years and older) with breast cancer who are human epidermal growth factor receptor 2 (HER2)/neu negative who are not administered HER2-targeted therapies

**Numerator Statement:** HER2-targeted therapies not administered during the initial course of treatment.

**Denominator Statement:** Adult women with breast cancer that are HER2 negative or HER2 undocumented.

**Exclusions:** Patient transfer to practice during or after initial course.

**Adjustment/Stratification:**

**Level of Analysis:** Clinician : Group/Practice

**Setting of Care:** Ambulatory Care : Clinician Office/Clinic

**Type of Measure:** Process

**Data Source:** Electronic Clinical Data : Registry

**Measure Steward:** American Society of Clinical Oncology

### STANDING COMMITTEE MEETING [05/18-19/2016]

#### 1. Importance to Measure and Report: The measure did not meet the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Previous Evidence Evaluation Accepted**; 1b. Performance Gap: **H-0; M-1; L-19; I-0**

##### Rationale:

- For the 2012 endorsement evaluation, the developer provided clinical practice guidelines from the American Society of Clinical Oncology (ASCO) and the Cancer Care Ontario (CCO) recommending trastuzumab to patients with HER2-positive node or node-negative breast cancer.
- For the current evaluation, the developer provided updated clinical practice guidelines from ASCO and CCO and an additional joint guideline from ASCO and the College of American Pathologists recommending HER2-targeted therapy for only for patients with HER2-positive breast cancer.
- The Committee agreed that the evidence is sufficient and there was no need to repeat the discussion and vote on evidence.
- The developer provided performance rates from the ASCO Quality Oncology Practice Initiative (QOPI®) Registry from 2013 – 2015. The mean performance rate in 2013 was 99.25%, 99.26% in 2014, and 99.54% in 2015. The developer provided 2013-2015 data stratified by race and/or ethnic groups that demonstrated little variation. Performance rates for Hispanics were 99.26% - 100.0% and 98.47% - 99.66% for black patients. The developer did not provide additional data on disparities.
- The Committee discussed the same issues related to performance gap that were discussed for [#1878](#).
- Like [#1878](#), the measure did not pass performance gap. Despite the high rate of performance other disparities may exist; therefore, the Committee voted to continue reviewing the measure against the rest of the criteria with the possibility of recommending the measure for inactive endorsement with reserve status.

#### 2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Previous Reliability Evaluation Accepted** 2b. Validity: **Previous Validity Evaluation Accepted**

##### Rationale:

### 1857 HER2 negative or undocumented breast cancer patients spared treatment with HER2-targeted therapies

- For the 2012 endorsement evaluation, data element validity testing was performed and counted for data element reliability. The dataset included 264 patient records from 44 QOPI practices submitted in spring 2007. Trained, independent nurse abstractors served as the “gold standard” against which practice abstractions were compared for accuracy. Kappa statistics were used to analyze the validity of the audited patient records compared to the submitted patient records. By convention, a kappa > 0.70 is considered acceptable. The developer provided a kappa score of 0.74 and an overall percent agreement of 96.0%. While this kappa score is above what is considered acceptable, the developer did not state which of the data elements this kappa score represented; no additional results were provided. NQF guidance states that testing should be done for all critical data elements. Like [#1878](#), the developer responded that they were unable to find the additional data from the testing previously conducted but based on the kappa score and overall agreement rate they did not have any concerns with the performance of the measure in the registry.
- Although the developer did not provide updated reliability and validity testing, the Committee agreed that the measure specifications were consistently implemented within the registry and accepted the previous reliability and validity evaluation.

#### 3. Feasibility: H-0; M-20; L-0; I-0

*(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)*

##### Rationale:

- This measure is based on clinical registry data and all data elements are available in electronic sources. The Committee agreed the measure is feasible.
- The Committee noted that eventual use of this measure through EHRs would lessen the data collection burden.

#### 4. Usability and Use: H-5; M-11; L-4; I-0

*(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)*

##### Rationale:

- The developer shared that this measure was recently selected for inclusion in a Medical Oncology Core Measure Set supported by AHIP and CMS. The measure was also recently approved for use in the Medicare Access & CHIP Reauthorization Act’s (MACRA) Merit-Based Incentive Payment System (MIPS).
- The Committee noted that this measure is used in Quality Oncology Practice Initiative (QOPI®).

#### 5. Related and Competing Measures

- No related or competing measures noted.

**Standing Committee Recommendation for Inactive Endorsement with Reserve Status : Y-19; N-1**

#### 6. Public and Member Comment

- 

**7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X**

**8. Board of Directors Vote: Y-X; N-X**

**9. Appeals**

## Measures Not Recommended

| 0459 Risk-Adjusted Length of Stay >14 Days after Elective Lobectomy for Lung Cancer   |
|---|
| <a href="#">Submission</a>   <a href="#">Specifications</a>   |
| <p><b>Description:</b> Percentage of patients aged 18 years and older undergoing elective lobectomy for lung cancer who had a prolonged length of stay &gt;14 days</p> <p><b>Numerator Statement:</b> Number of patients aged 18 years and older undergoing elective lobectomy for lung cancer who had a prolonged length of stay &gt;14 days</p> <p><b>Denominator Statement:</b> Number of patients aged 18 years and older undergoing elective lobectomy for lung cancer</p> <p><b>Exclusions:</b> None</p> <p><b>Adjustment/Stratification:</b></p> <p><b>Level of Analysis:</b> Facility, Clinician : Group/Practice</p> <p><b>Setting of Care:</b> Hospital/Acute Care Facility</p> <p><b>Type of Measure:</b> Outcome</p> <p><b>Data Source:</b> Electronic Clinical Data : Registry</p> <p><b>Measure Steward:</b> The Society of Thoracic Surgeons</p>   |
| <p><b>STANDING COMMITTEE MEETING [05/18-19/2016]</b></p> <p><b>1. Importance to Measure and Report: The measure did not meet the Importance Criteria</b><br/>(1a. Evidence, 1b. Performance Gap)</p> <p>1a. Evidence: <b>Previous Evidence Evaluation Accepted</b>; Initial 1b. Performance Gap: <b>H-0; M-10; L-5; I-6</b><br/><u>Re-vote on 1b. Performance Gap: M-11; L-8; I-1</u></p> <p><b>Rationale:</b></p> <ul style="list-style-type: none"> <li>For the 2008 endorsement evaluation,<sup>11</sup> the developers stated that prolonged length of stay after pulmonary lobectomy is both a surrogate marker of morbidity and a direct marker of increased resource utilization. Lower performing thoracic programs have the opportunity to design quality improvement initiatives when they know their rate of risk adjusted prolonged length of stay. The Committee accepted the previous evidence evaluation.</li> <li>For the current evaluation, the developers did not provide data on disparities from the measure as specified, but the Committee noted that there are studies demonstrating disparities based on the size of the program, the number of operations performed per year, insurance status, and general surgeons vs. board-certified thoracic surgeons.</li> <li>For the current evaluation, the developer provided performance data from the STS General Thoracic Surgery Database (GTSD) for patients that underwent elective lobectomy for lung cancer between July 1, 2012 and June 30, 2015 that demonstrated a mean prolonged length of stay (PLOS) (&gt;14 days) occurred in 4.3% of eligible patients. After the workgroup call the developer calculated the overall mean and median PLOS from 2009-2012, 2010-2013, 2011-2014, and 2012-2015. The PLOS decreased from a mean of 5.1% to 4.3% and the median decreased from 4.9% to 4.2% for all four time intervals. The Committee questioned whether 14 days was still an appropriate threshold for defining PLOS since LOS can be significantly impacted by surgical approach such as an open thoracotomy or a minimally-invasive thoracotomy as indicated in Wright et al 2010.</li> <li>The Committee noted that the number of patients per region ranged from 2,996 per 40 surgeons to 7,756 patients per 73 surgeons, yet the mean PLOS was ~4.0% for each region. The Committee was concerned that low-volume providers may affect overall performance rates making it difficult to distinguish high-performers from low-performers and determining if a gap in care exists based on the data provided.</li> <li>The Committee noted several concerns with the performance data provided by the developer.</li> </ul> |

### 0459 Risk-Adjusted Length of Stay >14 Days after Elective Lobectomy for Lung Cancer

The Committee requested that the developer provide performance data on 10 days vs. 14 days PLOS and the correlation between the number of procedures performed (volume) and PLOS at the next maintenance review of the measure.

- The Committee discussed the measure during the post-comment call and re-voted on the performance gap subcriterion. The Committee determined that the data provided did not demonstrate a gap in performance.

#### **2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria**

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Previous Reliability Evaluation Accepted** 2b. Validity: **M-15; L-6; I-0**

##### Rationale:

- For the 2012 endorsement evaluation,<sup>12</sup> the developer assessed test-retest reliability by comparing the results of estimated hospital rates of prolonged stay between 2 consecutive 6-month time intervals during 2009. The Pearson correlation between hospital-specific rates of prolonged stay in the first versus second half of 2009 was 0.31, which Evans (1996) suggests as “very weak” (0.20-0.39).
- For the current evaluation, the developer provided updated reliability testing of the measure score using the Pearson correlation coefficient to assess the signal-to-noise ratio. The reliability of the measure score increased as the volume of minimum procedures per year for participants increased. The reliability for all 244 participants in the registry and 23,174 operations was 37.6%, 44.5% for ≥10 procedures per year, and 63.8% for ≥ 40 procedures per year.
- The Committee agreed that the reliability scores of the measure were sufficient and accepted the previous reliability evaluation.
- For the 2012 endorsement evaluation,<sup>13</sup> the developer assessed face validity by an expert panel of thoracic surgeons assembled by the STS General Thoracic Surgery Database Task Force, the STS Task Force on Quality Initiatives and the STS Workforce on National Databases. The developer also stated that in 2010 they would conduct patient-level data element validity testing.
- For the current evaluation, the developer conducted data element validity testing using 10% of randomly selected STS GTSD participants from 2013 to 2015. Twenty cases (at least 15 lobectomy and up to 5 esophagectomy) that were previously submitted to the STS data warehouse were re-abstracted and compared to the “gold standard”. Agreement rates for the individual data elements ranged from 84.15% (diabetes control) to 100.0% (esophageal cancer, date of surgery, gastric outlet, and discharge date). The Committee agreed that the threats to validity were adequately assessed including the variables used in the risk-adjustment model. The Committee also agreed with the developer’s rationale that given the lack of consistent, compelling evidence regarding sociodemographic (SDS) factors and length of stay, there is no conceptual basis for adjusting the measure for SDS factors at this time, but noted that it is an important future state of development.

#### **3. Feasibility: H-17; M-4; L-0; I-0**

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)

##### Rationale:

- The Committee did not note any concerns regarding feasibility, acknowledging that some but not all of the data elements used to construct this measure are in defined fields in electronic sources.

#### **4. Usability and Use: H-8; M-11; L-2; I-0**

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

##### Rationale:

|  |
|--|
| <b>0459 Risk-Adjusted Length of Stay &gt;14 Days after Elective Lobectomy for Lung Cancer</b>  |
| <ul style="list-style-type: none"> <li>The measure is currently used for quality improvement by the STS General Thoracic Surgery Database which includes 273 participants. STS is planning to launch the general thoracic surgery component of STS Public Reporting Online in 2017.</li> <li>The developer did not provide any unintended consequences but the developer confirmed that if a patient was discharged to a LTAC (long-term acute care) facility on a ventilator on day 13 they would meet the measure.</li> </ul>  |
| <b>5. Related and Competing Measures</b>   |
| <ul style="list-style-type: none"> <li>No related or competing measures noted.</li> </ul>  |
| <b>Standing Committee Recommendation for Endorsement: Y-17; N-4</b>  |
| <b>6. Public and Member Comment</b>  |
| <ul style="list-style-type: none"> <li>One commenter suggest a measure addressing the discharge outcomes may provide better insight into variations of care due to low patient volume in the current measure. The commenter also noted the new measure(s) might be similar to measure #0460 with a different surgical procedure/patient diagnostic group.</li> <li>The developer response: Although length of stay is a surrogate for morbidity, measure #0459 is intended to be used to measure health care resource utilization. #1790 Risk-Adjusted Morbidity and Mortality for Lung Resection for Lung Cancer, an outcomes measure also stewarded by STS addresses the commenter's suggestion. In addition, STS recently developed a two-domain, outcomes only composite measure for lobectomy for lung cancer.</li> </ul> |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>   |
| <b>8. Board of Directors Vote: Y-X; N-X</b>  |
| <b>9. Appeals</b>  |

## 0460 Risk-Adjusted Morbidity and Mortality for Esophagectomy for Cancer

[Submission](#) | [Specifications](#)

**Description:** Percentage of patients aged 18 years and older undergoing elective esophagectomy for esophageal cancer who developed any of the following postoperative conditions: bleeding requiring reoperation, anastomosis leak requiring medical or surgical treatment, reintubation, ventilation >48 hours, pneumonia, or discharge mortality

**Numerator Statement:** Number of patients aged 18 years and older undergoing elective esophagectomy for esophageal cancer who developed any of the following postoperative conditions: bleeding requiring reoperation, anastomosis leak requiring medical or surgical treatment, reintubation, ventilation >48 hours, pneumonia, or discharge mortality.

**Denominator Statement:** Number of patients aged 18 years and older undergoing elective esophagectomy for esophageal cancer

**Exclusions:** None

**Adjustment/Stratification:**

**Level of Analysis:** Facility, Clinician : Group/Practice

**Setting of Care:** Hospital/Acute Care Facility

**Type of Measure:** Outcome

**Data Source:** Electronic Clinical Data : Registry

**Measure Steward:** The Society of Thoracic Surgeons

### STANDING COMMITTEE MEETING [05/18-19/2016]

#### 1. Importance to Measure and Report: The measure meets the Importance criteria

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Previous Evidence Evaluation Accepted**; 1b. Performance Gap: **H-6; M-12; L-2; I-1**

#### Rationale:

- In the 2008 endorsement evaluation,<sup>14</sup> the developer stated that measuring risk adjusted morbidity and mortality of patients undergoing esophagectomy for cancer provides surgeons and institutions the opportunity to evaluate outcomes and subsequently design quality improvement initiatives to address identified deficits. The Committee accepted the previous evidence evaluation.
- For the current evaluation, the developers did not provide data on disparities from the measure as specified. However, an analysis (Sammon et al, 2015) cited by the developer that was used to select patient factors for the risk model, suggested that age, gender, and race are relevant to esophagectomy outcomes. The Committee noted that race (African-Americans) was one of the variables included in the in the risk-model, therefore, taking into account race when computing the performance measure score.
- For the current evaluation, the developer provided performance data from the STS General Thoracic Surgery Database (GTSD) for patients that underwent elective esophagectomy for primary esophageal cancer between July 1, 2012 and June 30, 2015. The Committee noted that the median ranged from 27.7% to 28.6% and the 10<sup>th</sup> percentile and the 90<sup>th</sup> percentile ranged from 20.6% to 42.6%. The Committee agreed there is opportunity for improvement in care for patients undergoing elective esophagectomy.

#### 2. Scientific Acceptability of Measure Properties: The measure does not meet the Scientific Acceptability

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

Initial 2a. Reliability: **H-0; M-11; L-9; I-1** 2b. Validity: **M-12; L-9; I-0**

**Revote on 2a. Reliability: H-2; M-9; L-8; I-1** 2b. Validity: **M-10; L-9; I-1**

#### Rationale:

#### 0460 Risk-Adjusted Morbidity and Mortality for Esophagectomy for Cancer

- For the 2012 endorsement evaluation,<sup>15</sup> the developer assessed test-retest reliability by comparing the results of estimated hospital rates mortality or major morbidity between 2 consecutive 6-month time intervals during 2009. The Pearson correlation between hospital-specific rates of mortality or major morbidity in the first versus second half of 2009 was 0.50, which Evans (1996) suggests as “weak” (0.40-0.59).
- For the current evaluation, the developer provided updated reliability testing of the measure score using the Pearson correlation coefficient to assess the signal-to-noise ratio. The reliability of the measure score increased as the volume of minimum procedures per year for participants increased. The reliability scores for all 169 participants and 4,557 operations were 44.4%, 67.9% for ≥5 procedures per year, and 80.6% for ≥20 procedures per year. The Committee noted that more than 55.0% of participants (94) in the registry did fewer than 5 procedures a year. The Committee expressed their concerns with the reliability of this low-volume procedure and that the measure was not specified for ≥5 procedures per year. The Committee also expressed their concerns with combining morbidity and mortality and asked the developer if there were plans for differential weighting of these outcomes. The developer responded that they were developing a new measure that more heavily weights mortality than morbidity and it would be complete by the next maintenance review. The previous Committee also noted the same concerns in 2012.
- Due to the concerns regarding the reliability of the measure as specified, the Committee and did not reach consensus on this criterion.
- For the 2012 endorsement evaluation,<sup>16</sup> the developer assessed face validity by an expert panel of thoracic surgeons assembled by the STS General Thoracic Surgery Database Task Force, the STS Task Force on Quality Initiatives and the STS Workforce on National Databases. The developer also stated that in 2010 they would conduct patient-level data element validity testing.
- For the current evaluation, the developer conducted data element validity testing using 10% of randomly selected STS GTSD participants from 2013 to 2015. Twenty cases (at least 15 lobectomy and up to 5 esophagectomy) that were previously submitted to the STS data warehouse were re-abstracted and compared to the “gold standard”. Agreement rates for the individual data elements ranged from 84.15% (diabetes control) to 100.0% (esophageal cancer, date of surgery, gastric outlet, and discharge date). The Committee agreed that the risk-model variables were appropriate.
- The Committee determined that the data element validity testing was adequate but the data provided demonstrated a threat to validity due to low-volume providers. On re-vote the Committee did not pass the reliability and validity subcriteria.

#### 3. Feasibility: H-9; M-12; L-0; I-0

*(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/unintended consequences identified 3d. Data collection strategy can be implemented)*

##### Rationale:

- The Committee did not note any concerns regarding feasibility, acknowledging that some but not all of the data elements used to construct this measure are in defined fields in electronic sources.

#### 4. Usability and Use: H-4; M-15; L-2; I-0

*(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)*

##### Rationale:

- The measure is currently used for quality improvement by the STS General Thoracic Surgery Database which includes 273 participants. STS is planning to launch the general thoracic surgery component of STS Public Reporting Online in 2017.
- The Committee noted that it would be important to determine how to publicly report the performance rates of this measure for the layperson (i.e. low-volume versus low performance).

|  |
|--|
| <b>0460 Risk-Adjusted Morbidity and Mortality for Esophagectomy for Cancer</b>   |
| <b>5. Related and Competing Measures</b> <ul style="list-style-type: none"> <li>No related or competing measures noted.</li> </ul> |
| <b>Standing Committee Recommendation for Endorsement: Y-20; N-1</b>  |
| <b>6. Public and Member Comment</b> <ul style="list-style-type: none"> <li></li> </ul>   |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>   |
| <b>8. Board of Directors Vote: Y-X; N-X</b>  |
| <b>9. Appeals</b>  |

## 2936 Admissions and Emergency Department (ED) Visits for Patients Receiving Outpatient Chemotherapy

### [Submission](#) |

**Description:** Measure estimates hospital-level, risk-adjusted rates of inpatient admissions or ED visits for cancer patients >18 years of age for at least one of the following diagnoses—anemia, dehydration, diarrhea, emesis, fever, nausea, neutropenia, pain, pneumonia, or sepsis—within 30 days of hospital outpatient chemotherapy treatment. The two rates are calculated and reported separately.

**Numerator Statement:** This measure involves calculating two mutually exclusive outcomes: one or more inpatient admissions or one or more ED visits for any of the following diagnoses—anemia, dehydration, diarrhea, emesis, fever, nausea, neutropenia, pain, pneumonia, or sepsis—within 30 days of chemotherapy treatment among cancer patients receiving treatment in a hospital outpatient setting. These 10 conditions are potentially preventable through appropriately managed outpatient care. The qualifying diagnosis on the admission or ED visit claim must be (1) the principal diagnosis or (2) a secondary diagnosis accompanied by a principal diagnosis of cancer.

**Denominator Statement:** The measure cohort includes Medicare FFS patients aged 18 years and older as of the start of the performance period with a diagnosis of any cancer who received at least one hospital outpatient chemotherapy treatment at the reporting hospital during the performance period.

**Exclusions:** We established the following exclusion criteria after reviewing the literature, examining existing measures, reviewing feedback from a public comment period, and discussing alternatives with the Cancer Working Group and TEP members (see Section Ad.1. for description of group and membership). The goal was to be as inclusive as possible; we excluded only those patient groups for which hospital visits were not typically a quality signal or for which risk adjustment would not be adequate. The exclusions, based on clinical rationales, prevent unfair distortion of performance results.

1) Patients with a diagnosis of leukemia at any time during the performance period.

Rationale: Patients with leukemia are excluded due to the high toxicity of treatment and recurrence of disease so that admissions do not reflect poorly managed outpatient care for this population. Patients with leukemia have an expected admission rate due to relapse, so including leukemia patients in the cohort could be conceptualized as a planned admission, which does not align with the intent of the measure.

2) Patients who were not enrolled in Medicare FFS Parts A and B in the year prior to the first outpatient chemotherapy treatment during the performance period.

Rationale: We exclude these patients to ensure complete patient diagnosis data for the risk-adjustment model, which uses the year prior to the first chemotherapy treatment during the period to identify comorbidities.

3) Patients who do not have at least one outpatient chemotherapy treatment followed by continuous enrollment in Medicare FFS Parts A and B in the 30 days after the procedure.

Rationale: We exclude these patients to ensure full data availability for outcome assessment.

### **Adjustment/Stratification:**

**Level of Analysis:** Facility

**Setting of Care:** Hospital/Acute Care Facility

**Type of Measure:** Outcome

**Data Source:** Administrative claims

**Measure Steward:** Centers for Medicare & Medicaid Services (CMS)

### **STANDING COMMITTEE MEETING [05/18-19/2016]**

#### **1. Importance to Measure and Report: Consensus was not reached on the Importance criteria**

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-12; N-9**; 1b. Performance Gap: **H-2; M-9; L-3; I-7**

### 2936 Admissions and Emergency Department (ED) Visits for Patients Receiving Outpatient Chemotherapy

#### Rationale:

- According to the developer, chemotherapy treatment can have severe, predictable side effects, and hospital admissions and ED visits among patients receiving treatment in a hospital outpatient setting are often caused by manageable side effects and complications. Admissions and ED visits for eligible diagnoses—anemia, dehydration, diarrhea, emesis, fever, nausea, neutropenia, pain, pneumonia, or sepsis—may be due to patients receiving treatment in a hospital outpatient setting having unmet needs and gaps in care, which, if addressed, could reduce admissions and ED visits and increase patients' quality of life. Treatment plans and guidelines exist to support the management of these conditions. Hospitals that provide outpatient chemotherapy should implement appropriate care to minimize the need for acute hospital care for these adverse events.
- The Committee acknowledged this outcome measure encourages care coordination and symptom management in an effort to minimize the side effects of chemotherapy administration, improve the quality of cancer care, and patients' overall quality of life. While the Committee agreed that the interventions outlined by the developer to prevent and manage anemia, dehydration, diarrhea, nausea/emesis, neutropenic fever, pain, and pneumonia/sepsis improves patients' quality of life, some members stated that evidence linking these interventions with decreased hospitalizations and ED visits was not provided. Committee members agreed there are higher levels of evidence to support some of the clinical interventions the developer listed versus non-clinical interventions like care coordination. Some Committee members expressed concern with the complexity of the measure and the broad range of diagnoses that it would be difficult for a facility to determine where to focus their quality improvement efforts. Other Committee members agreed that list of side effects are broad but determined that the developer provided sufficient evidence to support the interventions to prevent and manage the side effects and symptoms of chemotherapy and decrease the risk of ED visits and hospital admissions.
- The Committee did not reach consensus on the evidence of a linkage between the broad range of side effects and reduced ED visits and hospitalizations.
- The developer provided inpatient admission rates and ED visit rates from July 1, 2012 - June 30, 2013 using Medicare FFS claims for 252,408 patients and 3,765 hospitals. The risk-standardized inpatient admission rate ranged from 6.0% to 24.9% (median 10.2, 25<sup>th</sup> and 75<sup>th</sup> percentiles were 9.8 and 10.8, respectively). The risk-standardized ED visit rate ranged from 2.1% to 7.5% (median 4.1, 25<sup>th</sup> and 75<sup>th</sup> percentiles are 4.0 and 4.4, respectively). Additionally, the developer cited several studies that demonstrated a significant number of cancer patients experience inpatient admissions and ED visits each year related to the frequently reported side effects of chemotherapy. Other studies cited by the developer suggest that there is substantial institutional and geographic variation in hospital admissions and ED visits among chemotherapy patients.
- The developer did not provide disparities data from the measure as specified but did examine associations between outcomes and sociodemographic (SDS) factors. The developer analyzed dual-eligibility, race, and AHRQ SES Composite Index to determine if these factors affected whether patients receiving hospital-based outpatient chemotherapy were more likely to have an inpatient admission and emergency department visit within 30 days than "non-low SDS" patients. On the patient level, the developer's analysis found disparities based on the 3 variables examined. However, these disparities were no longer significant when evaluated at the hospital level. One of the Committee members noted that disparities may have not been significant at the hospital level due to volume or other statistical issues. The member suggested stratifying the measure since disparities were significant at the patient level.
- The Committee noted the narrow interquartile range (IQR) for both rates indicating little variability in performance among most of the facilities. On the other hand, the Committee noted

### 2936 Admissions and Emergency Department (ED) Visits for Patients Receiving Outpatient Chemotherapy

the overall range of inpatient admission rates demonstrated a gap in care and an opportunity for improvement, especially for the facilities in the 25<sup>th</sup> percentile. A Committee member questioned that sufficient data was provided to determine if a gap in care existed due to the significantly small difference in percentage points between the 25<sup>th</sup> and 75<sup>th</sup> percentile on the ED visit rate. The Committee asked the developer if they could provide the 10<sup>th</sup> and 90<sup>th</sup> percentile rates for the ED visit rates; the developer did not have the data available at the in-person meeting.

- The Committee did not reach consensus on the performance gap criterion.
- Prior to the Committee's vote on performance gap, NQF staff recommended that the Committee consider the measure as 2 separate measures and NQF would categorize it as a "paired measure" due to the developer's wish that the 2 rates be reported together. As stated in previous conversations with NQF staff, the developer expressed the intent of the measure to calculate 2 rates and report them separately. However, 90% of the Committee voted to keep the measure as it was submitted; and only 10% voted to separate the measure into a paired measure, so the measure was not split.

#### **2. Scientific Acceptability of Measure Properties: The measure did not meet the Scientific Acceptability criteria**

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-0; M-4; L-10; I-6** 2b. Validity: **N/A**

#### Rationale:

- The developer used 2012-2013 Medicare data from 3,765 hospitals and 240,446 patients. A total of 942 hospitals with ≥ 60 patients in the cohort were included in the sample. A split-sample methodology was used to test the measure score reliability. The developers randomly assigned half of the patients in each hospital to 2 separate groups, calculated the performance measure score for each hospital in each of the 2 groups, and calculated the Pearson correlation between the performance rates in each half-year sample; the higher the correlation, the higher the reliability of the measure.
- The developers also used the intraclass correlation coefficient (ICC) signal-to-noise method to determine the recommended minimum number of cases needed to maintain a reliability level of 0.4 or higher. The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospitals.
- The reliability score for inpatient admissions was 0.41 and 0.27 for ED visits. To achieve reliability (ICC) of 0.4, a minimum of 25 patients are required to calculate the inpatient admissions rate and a minimum of 20 patients for the ED visit rate per performance period. The developer recommended a performance period long enough to accumulate a sufficient number of patients per hospital for improved reliability.
- During the workgroup call and the in-person meeting, the Committee questioned the developer about the strength of the reliability score for the ED measure (Pearson correlation = 0.27). The developer responded that they had access to only 1 year of data at the time the analyses were conducted. As mentioned previously, the methodology requires a random-split of data into 2 distinct samples to calculate a test-retest reliability score. Therefore, the test-retest reliability calculation was based on correlation between 2 half-year samples, or roughly half the data that will be used to calculate outcome rates for public reporting. Calculating reliability estimates on samples analogous in size to those in public reporting would require 2 years of data, to which developers currently do not have access, but is expected to increase the measure reliability. The developer also noted that given the reliability calculation split a year of data into 2 half-year samples, they were further limited by low facility volume and the low rate of the ED measure (median rate of 4.1 per 100 patient visits for hospital-based outpatient chemotherapy). The

### 2936 Admissions and Emergency Department (ED) Visits for Patients Receiving Outpatient Chemotherapy

reliability score measures the consistency within two split samples, and diminishing sample size in the presence of low rates makes it less likely that the two split samples are similar. For example, a facility with 30 cases would be expected to have approximately one ED visit case. When this facility is randomly split the observed event would only be attributed to 1 of the 2 split-year samples, resulting in a discrepancy in the rates of 0% versus 6.7%. These large discrepancies reduce the strength of the reliability estimate.

- To determine what effect a 2-year sample of data would have on reliability estimates, the developers conducted additional reliability analyses with the same split-year samples used in the original analyses. Specifically, the developers recalculated the reliability score using the Intraclass Correlation Coefficient (ICC) and adjusted the ICC estimate using the Spearman-Brown prophecy formula to determine the range of ICC values if calculated in a split-sample with 2 years of data instead of 1. The Spearman-Brown prophecy formula provides an estimate of an ICC if the number of items in a test increases by a certain factor. Using the Spearman-Brown formula, assuming a 2-year split sample, the adjusted ICC was estimated to be 0.47 (95% confidence interval: 0.40/0.53). Accordingly, the developers expect reliability estimates will improve when calculated in a data sample analogous in size to that used for implementation and public reporting.
- During the workgroup call, one of the Committee members inquired about the distribution of hospital case count, since the number of hospitals included in the reliability analysis declined from over 3,000 to less than 1,000 once the minimum patient threshold was imposed on the split-halves analysis. The developer conducted additional analyses and found that 41.0% of hospitals had a minimum case count of  $\geq 25$  patients (the typical threshold for public reporting) over the 1 year period from July 2012 through June 2013.
- Some Committee members continued to express their concern with the complexity of the measure and questioned a facility's ability to consistently implement the measure and the potential impact on reliability. The developer clarified that the measure does not require facilities to calculate their rates; rather the rates are calculated by CMS using Medicare FFS administrative claims.
- Another Committee member expressed their concerns with the numerator limiting admissions/rates to inpatient and ED. Many facilities and cancer centers, the member reasoned, have affiliated urgent care centers or 24-hour clinics rather than emergency departments. If a patient was seen at an urgent care centers or clinics for one of the eligible diagnoses, they would not be counted in the numerator. Additionally, if they were admitted to the hospital for observation, they would not be included in the numerator unless they crossed the two-midnight rule.<sup>17</sup>
- Committee members also voiced their concerns about attribution. One member suggested that if a patient receives chemotherapy from more than 1 facility in the 30 day timeframe, the facility that administered the chemotherapy prior to the inpatient admission or ED visit should bear more attribution. The developer pointed to the analysis they conducted to see how many patients received chemotherapy from more than 1 hospital and found that only 5% of patients in the sample (n=240,446) received chemotherapy at more than 1 hospital.
- Other Committee members noted that patients receiving concurrent chemoradiotherapy should be excluded from the denominator.
- Overall, the Committee concluded that the measure did not meet the reliability criterion due to the concerns discussed, specifically the small sample size used for reliability testing and the low reliability scores.

### 3. Feasibility: N/A

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/

|  |
|--|
| <b>2936 Admissions and Emergency Department (ED) Visits for Patients Receiving Outpatient Chemotherapy</b>   |
| <i>unintended consequences identified 3d. Data collection strategy can be implemented)</i><br><u>Rationale:</u><br>•   |
| <b>4. Usability and Use: N/A</b><br><i>(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)</i><br><u>Rationale:</u><br>• |
| <b>5. Related and Competing Measures</b><br>• No related or competing measures noted.  |
| <b>Standing Committee Recommendation for Endorsement: N/A</b><br><u>Rationale</u><br>• The measure did not pass reliability.   |
| <b>6. Public and Member Comment</b><br>•   |
| <b>7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X</b>   |
| <b>8. Board of Directors Vote: Y-X; N-X</b>  |
| <b>9. Appeals</b>  |

<sup>1</sup> American Cancer Society. Economic Impact of Cancer. <http://www.cancer.org/cancer/cancerbasics/economic-impact-of-cancer>. Published February 6, 2015. Accessed June 27, 2016.

<sup>2</sup> American Cancer Society. 2016 Estimated Deaths for the Four Major Cancers by Sex and Age Group. Cancer Facts & Figures 2016. <http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2016/index>. Accessed June 27, 2016.

<sup>3</sup> American Cancer Society. Cancer Facts & Figures 2016. <http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2016/index>. Accessed June 27, 2016.

<sup>4</sup> National Cancer Institute. National Expenditure for Cancer Sites. Cancer Prevalence and Cost of Care Projections. <https://costprojections.cancer.gov/expenditures.html>. Accessed June 28, 2016.

<sup>5</sup> Measure #0508 was initially endorsed in 2008 with time-limited endorsement.

<sup>6</sup> In 2012, #0508 underwent time-limited reliability testing review and received full endorsement.

<sup>7</sup> In 2012, #0508 underwent time-limited validity testing review and received full endorsement.

<sup>8</sup> Measure #0509 was initially endorsed in 2008 with time-limited endorsement.

<sup>9</sup> In 2012, #0509 underwent time-limited reliability testing review and received full endorsement.

<sup>10</sup> In 2012, #0509 underwent time-limited validity testing review and received full endorsement.

<sup>11</sup> Measure #0459 was initially endorsed in 2008 with time-limited endorsement.

<sup>12</sup> In 2012, #0459 underwent time-limited reliability testing review and received full endorsement.

<sup>13</sup> In 2012, #0459 underwent time-limited validity testing review and received full endorsement.

<sup>14</sup> Measure #0460 was initially endorsed in 2008 with time-limited endorsement.

<sup>15</sup> In 2012, #0460 underwent time-limited reliability testing review and received full endorsement.

<sup>16</sup> In 2012, #0460 underwent time-limited validity testing review and received full endorsement.

<sup>17</sup> CMS announced the two-midnight rule in 2013. Under this rule, only patients that the physician expects will need to spend two nights in the hospital would be considered as hospital inpatients. [http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief\\_id=133](http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=133)