NATIONAL QUALITY FORUM
Driving measurable health improvements together

Memo

October 5, 2020

To: Cancer Standing Committee

From: NQF staff

Re: Post-comment web meeting to discuss public comments received and NQF member expression of support

Introduction

NQF closed the public commenting period on the measures submitted for endorsement consideration. NQF received two comments that will be considered by the Standing Committee.

Purpose of the Call

The Cancer Standing Committee will meet via web meeting on October 5, 2020 from 12:00-2:00pm ET. The purpose of this call is to:

- Review and discuss comments received during the post-evaluation public and member comment period;
- Provide input on proposed responses to the post-evaluation comments;
- Review and discuss NQF members' expression of support of the measures under consideration; and
- Determine whether reconsideration of any measures or other courses of action are warranted.

Standing Committee Actions

1. Review this briefing memo and draft report.
2. Review and consider the full text of all comments received and the proposed responses to the post-evaluation comments (see comment table and additional documents included with the call materials).
3. Review the NQF members' expressions of support of the submitted measures.
4. Be prepared to provide feedback and input on proposed post-evaluation comment responses.

Conference Call Information

Please use the following information to access the conference call line and webinar:

Speaker dial-in #: **1-800-768-2983**, Access Code: **9037531**
Web link: https://core.callinfo.com/callme/?ap=8007682983&ac=9037531&role=p&mode=ad

Background

Cancer is the second most common cause of death in the U.S., exceeded only by heart disease. NCI estimated that in 2018, 1.7 million new cases of cancer would be diagnosed in the United States and over 600,000 people will die from the disease. Furthermore, nearly half of all men and one-third of all women in the U.S. will develop cancer during their lifetime. In addition, diagnosis and treatment of

cancer has great economic impact on patients, their families, and society. NCI estimated that, in 2010, the costs for cancer care in the U.S. totaled nearly $157 billion and could reach $174 billion in 2020.

Cancer care is complex and provided in multiple settings—hospitals, outpatient clinics, ambulatory infusion centers, radiation oncology treatment centers, radiology departments, palliative and hospice care facilities—and by multiple providers including surgeons, oncologists, nurses, pain management specialists, pharmacists and social workers.

The Cancer Standing Committee oversees NQF's portfolio of Cancer measures that includes measures for hematology, breast cancer, colon cancer, prostate cancer, and other cancer measures. The purpose of this project was to review Cancer measure submitted for endorsement or undergoing maintenance during the Spring 2020 cycle.

During the Measure Evaluation Web Meeting held on July 10, 2020, the Cancer Standing Committee evaluated one maintenance measure for endorsement consideration. The measure failed to pass on validity.

Comments Received

NQF solicits comments on measures undergoing review in various ways and at various times throughout the evaluation process. First, NQF solicits comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). Second, NQF solicits member and public comments during a 16-week comment period via an online tool on the project webpage.

Pre-evaluation Comments

NQF solicits comments prior to the evaluation of the measures via an online tool on the project webpage. For this evaluation cycle, the pre-evaluation comment period was open from May 11 to June 19, 2020 for the measure under review. No public or NQF comments were received.

Post-evaluation Comments

The draft report was posted on the project webpage for public and NQF member comment on August 14, 2020 for 30 calendar days. During this commenting period, NQF received two comments from two member organizations:

Member Council	# of Member Organizations Who Commented
Health Professional	2

We have included all comments that we received (both pre- and post-evaluation) in the comment table (excel spreadsheet) posted to the Committee SharePoint site. This comment table contains the commenter's name, comment, associated measure, topic (if applicable), and—for the post-evaluation comments— responses (including measure steward/developer responses) for the Committee's consideration. Please review this table in advance of the meeting and consider the individual comments received and the proposed responses. The Standing Committee's recommendations will be reviewed by the Consensus Standards Approval Committee (CSAC) on November 17-18, 2020. The CSAC will determine whether or not to uphold the Standing Committee's recommendation for each measure submitted for endorsement consideration. All committee members are encouraged to attend the CSAC meeting to listen to the discussion.

Please note measure stewards/developers were asked to respond where appropriate. Where possible, NQF staff has proposed draft responses for the Committee to consider.

Comments and Their Disposition

Measure-Specific Comments

508: Radiology: Inappropriate Use of a "Probably Benign" Assessment Category in Screening

The American Geriatrics Society (AGS) wishes to provide comment on NQF# 0508 Diagnostic Imaging: Inappropriate Use of "Probably Benign" Assessment Category in Screening Mammograms. We note that this measure as written is somewhat confusing. We are unclear if the intent is that screening mammograms should never be labeled as "probably benign" and if in this gray zone should always be further confirmed (by comparing to prior mammograms for example) as benign or not benign. The wording of "inappropriate use of probably benign assessment" was not clear - is there an appropriate use for this category? We also think that the measure should specify this is only for radiologists reading mammograms; we do not see how this would apply to other clinicians.

Measure Steward/Developer Response:

The "probably benign" assessment category (or Breast Imaging Reporting and Data System (BI-RADS) 3) is reserved for findings with a high probability (≥98%) of being benign and should not be used as a category for indeterminate findings. Inappropriate designation of findings as "probably benign" can result in the unnecessary follow-up of lesions that could be quickly classified or delayed diagnosis and treatment of cancerous lesions. It is of further note that NQF #508 guidance documents, specifically BIRADS 3, was commented on during the September 14, 2020, NQF Reducing Diagnostic Error Project meeting. The comment specifically focused on the importance of appropriateness measures improving diagnostic accuracy.

During the Standing Committee meeting convened on July 10, 2020, ACR noticed inconsistencies regarding the Standing Committee's interpretation of the measure as it applies to implementation and measurement elements. Particularly, the Standing Committee's recommendation regarding re-specifying the measure to capture "follow-up mammograms vs first-time mammograms only", rather than the current specifications. ACR respectfully disagrees that the measure should be re-specified for "follow-up" mammograms only. It appears there was not a clear understanding as to the distinction between a screening mammogram and a diagnostic mammogram. While screening mammograms are routinely administered at regular intervals to detect breast cancer in asymptomatic patients, diagnostic mammograms are used after abnormal or suspicious results on a screening mammogram or after some signs of breast cancer alert the physician to check the tissue. Regular interval screening mammograms after an initial or baseline exam are not technically considered "follow-up" exams, thus the recommendation to only capture "follow-up" exams in the measure denominator would be excluding patients with baseline screening mammograms for whom it would also not be advantageous to recommend.

The current measure evaluates assessment of findings on annual or bi-annual, regular interval screening mammography exams as to whether abnormal or suspicious findings are followed up efficiently and expeditiously. Because a "probably benign" (BI-RADS 3) assessment at screening defers the diagnostic workup by six months, it is strongly recommended that BI-RADS 3 assessments are issued only on a diagnostic mammography exam after an appropriate workup. Thus, to only capture regular interval screening exams in the measure denominator would be excluding patients with baseline or "first-time" screening mammograms for whom it would also

be disadvantageous to defer diagnostic workup. This recommendation has been made based on the following studies which indicate the major advantages that full diagnostic imaging evaluation will provide, in addition to identifying both benign and malignant lesions promptly instead of waiting for six months.

1. More prompt identification of truly benign findings (simple cysts, some intramammary lymph nodes, some cases of grouped skin calcifications, etc.). A large-scale Breast Cancer Surveillance Consortium (BSCS) study, involving more than 1 million mammograms, has shown that recall imaging significantly increases the identification of characteristically benign lesions, thus promptly establishing a benign diagnosis, eliminating 6 months of potential anxiety, and obviating short-interval follow-up examination. (Kerlikowske K, Smith-Bindman R, Abraham LA, et al.

2. More prompt identification of some rapidly growing cancers (the same BCSC study also suggested that recall imaging leads to the prompt diagnosis of some aggressively growing cancers by identifying these tumors when they are smaller and more likely to be node-negative, rather than six months later at initial short-interval follow-up examination.)

NQF #508 involves reporting the percentage of screening (as opposed to diagnostic) mammography examinations that are assessed as BI-RADS category 3, with the stated goal of reducing this to "approaching 0%" in clinical practice, a BI-RADS category 3 assessment rendered from a screening exam, without prompt diagnostic workup, is considered a positive screening exam. The rationale for making BI-RADS category 3 at screening positive is that it recommends additional imaging evaluation prior to routine screening in one year. Use of BI-RADS category 3 assessment at screening is no longer a strategy to reduce recall rate

ACR also plans to determine the necessary data elements to identify disparities that demonstrate a larger performance gap and examine the performance variance among larger and smaller practices (e.g., low volume: 20 patients and high volume: 100 patients).

We acknowledge that NQF #508 did not meet NQF's must-pass criterion to achieve appropriate empirical validity evidence based on the testing data submitted, which hypothesized that physicians who perform well on NQF #508 would also perform well on the other related measures. Unfortunately, we did not find a strong correlation for performance between these measures using the construct validity method. However, ACR plans to re-assess the methodology appropriate for establishing validity and reanalyze the data collected for NQF #508 once the measure is updated appropriately, following potential revisions associated with the previously mentioned Standing-Committee feedback. Such specification updates and validity testing methodology could present a strong justification for this measure's endorsement.~~Developer responses will be added when available prior to the post-comment call scheduled on October 5, 2020.~~

Proposed Committee Response:

Thank you for your comments. The Committee will review these comments during its deliberations on the post-comment call scheduled on October 5, 2020.

Action Item:

The Committee should review the comments and the developer's response and be prepared to discuss any recommendations for the developer to consider.

The American College of Radiology (ACR) provided a comment to address a portion of the Standing Committee's feedback and ACR's intention to address the associated issues in the near future. Specifically, ACR commented on inconsistencies regarding the Standing Committee's interpretation of measure NQF 0508 as it applies to implementation and measurement elements. The ACR referenced the Standing Committee's recommendation regarding re-specifying the measure to capture "follow-up mammograms vs first-time mammograms only", rather than the current specifications. ACR respectfully disagrees that the measure should be re-specified for "follow-up" mammograms only.

The ACR further mentioned that they plan to determine the necessary data elements to identify disparities that demonstrate a larger performance gap and examine the performance variance among larger and smaller practices (e.g., low volume: 20 patients and high volume: 100 patients). The ACR also acknowledged that NQF #508 did not meet NQF's must-pass criterion to achieve appropriate empirical validity evidence based on the testing data submitted. The ACR plans to re-assess the methodology appropriate for establishing validity and reanalyze the data collected for NQF #508 once the measure is updated appropriately, following potential revisions associated with the previously mentioned Standing-Committee feedback. Please find the full comment provided in Appendix B.

Proposed Committee Response:

Thank you for your comments. The Committee will review these comments during its deliberations on the Post-Comment Call scheduled on October 5, 2020.

Action Item:

The Committee should review the comments and be prepared to discuss any recommendations for the developer to consider.

NQF Member Expression of Support

Throughout the 16-week continuous public commenting period, NQF members had the opportunity to express their support ("support" or "do not support") for each measure submitted for endorsement consideration to inform the Committee's recommendations. No NQF members provided their expressions of support: See Appendix A.

Appendix A: NQF Member Expression of Support Results

No NQF members provided their expressions of support.

Appendix B: NQF Member and Public Comments

Member and Public Full Comments

The American College of Radiology (ACR), measure developer and steward of NQF #508: Radiology: Inappropriate Use of a "Probably Benign" Assessment Category in Screening Mammograms, appreciates NQF's Cancer Standing Committee endorsement review. The following comments address a portion of the Standing Committee's feedback and ACR's intention to address the associated issues in the near future. Additionally, we emphasize that the "probably benign" assessment category (or Breast Imaging Reporting and Data System (BI-RADS) 3) is reserved for findings with a high probability (≥98%) of being benign and should not be used as a category for indeterminate findings. Inappropriate designation of findings as "probably benign" can result in the unnecessary follow-up of lesions that could be quickly classified or delayed diagnosis and treatment of cancerous lesions. It is of further note that NQF #508 guidance documents, specifically BIRADS 3, was commented on during the September 14, 2020, NQF Reducing Diagnostic Error Project meeting. The comment specifically focused on the importance of appropriateness measures improving diagnostic accuracy.

Several topics that rely on women's imaging experts' input were discussed during the Standing Committee's virtual meeting and summarized in the Draft CDP report. As such, the following addresses ACR's response to the Standing Committee's recommendations. During the Standing Committee meeting convened on July 10, 2020, ACR noticed inconsistencies regarding the Standing Committee's interpretation of the measure as it applies to implementation and measurement elements. Particularly, the Standing Committee's recommendation regarding re-specifying the measure to capture "follow-up mammograms vs first-time mammograms only", rather than the current specifications. ACR respectfully disagrees that the measure should be re-specified for "follow-up" mammograms only. It appears there was not a clear understanding as to the distinction between a screening mammogram and a diagnostic mammogram. While screening mammograms are routinely administered at regular intervals to detect breast cancer in asymptomatic patients, diagnostic mammograms are used after abnormal or suspicious results on a screening mammogram or after some signs of breast cancer alert the physician to check the tissue. Regular interval screening mammograms after an initial or baseline exam are not technically considered "follow-up" exams, thus the recommendation to only capture "follow-up" exams in the measure denominator would be excluding patients with baseline screening mammograms for whom it would also not be advantageous to recommend.

The current measure evaluates assessment of findings on annual or bi-annual, regular interval screening mammography exams as to whether abnormal or suspicious findings are followed up efficiently and expeditiously. Because a "probably benign" (BI-RADS 3) assessment at screening defers the diagnostic workup by six months, it is strongly recommended that BI-RADS 3 assessments are issued only on a diagnostic mammography exam after an appropriate workup. Thus, to only capture regular interval screening exams in the measure denominator would be excluding patients with baseline or "first-time" screening mammograms for whom it would also be disadvantageous to defer diagnostic workup. This recommendation has been made based on the following studies which indicate the major advantages that full diagnostic imaging evaluation will provide, in addition to identifying both benign and malignant lesions promptly instead of waiting for six months.

1. More prompt identification of truly benign findings (simple cysts, some intramammary lymph nodes, some cases of grouped skin calcifications, etc.). A large-scale Breast Cancer Surveillance Consortium (BSCS) study, involving more than 1 million mammograms, has shown that recall imaging significantly increases the identification of characteristically benign lesions, thus promptly establishing a benign diagnosis, eliminating 6 months of potential anxiety, and

obviating short-interval follow-up examination. (Kerlikowske K, Smith-Bindman R, Abraham LA, et al.

2. More prompt identification of some rapidly growing cancers (the same BCSC study also suggested that recall imaging leads to the prompt diagnosis of some aggressively growing cancers by identifying these tumors when they are smaller and more likely to be node-negative, rather than six months later at initial short-interval follow-up examination.)

NQF #508 involves reporting the percentage of screening (as opposed to diagnostic) mammography examinations that are assessed as BI-RADS category 3, with the stated goal of reducing this to "approaching 0%" in clinical practice, a BI-RADS category 3 assessment rendered from a screening exam, without prompt diagnostic workup, is considered a positive screening exam. The rationale for making BI-RADS category 3 at screening positive is that it recommends additional imaging evaluation prior to routine screening in one year. Use of BI-RADS category 3 assessment at screening is no longer a strategy to reduce recall rate.

ACR also plans to determine the necessary data elements to identify disparities that demonstrate a larger performance gap and examine the performance variance among larger and smaller practices (e.g., low volume: 20 patients and high volume: 100 patients).

We acknowledge that NQF #508 did not meet NQF's must-pass criterion to achieve appropriate empirical validity evidence based on the testing data submitted, which hypothesized that physicians who perform well on NQF #508 would also perform well on the other related measures. Unfortunately, we did not find a strong correlation for performance between these measures using the construct validity method. However, ACR plans to re-assess the methodology appropriate for establishing validity and reanalyze the data collected for NQF #508 once the measure is updated appropriately, following potential revisions associated with the previously mentioned Standing-Committee feedback. Such specification updates and validity testing methodology could present a strong justification for this measure's endorsement.