TO: Consensus Standards Approval Committee (CSAC)

FR: Kathryn Goodwin, Senior Project Manager, and Karen Johnson, Senior Director

RE: Musculoskeletal Off-Cycle Measure Review

DA: April 26, 2017

The CSAC will review recommendations from the Musculoskeletal Off-Cycle Measure Review project at its May 1, 2017 in-person meeting.

This memo includes a summary of the project, the measures under evaluation, and the public and member comments and the standing committee's responses to the comments.

Because the Standing Committee did not recommend the measures for endorsement, NQF did not release the measures for NQF Member voting.

Accompanying this memo are the following documents:

1. **Musculoskeletal: Off-Cycle Measure Review 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**. This table lists 10 comments received and the Standing Committee responses.

BACKGROUND

The off-cycle activities of the Musculoskeletal Committee in late 2016 and early 2017 have focused on the evaluation of two measures that were last considered in 2014. At that time, the Musculoskeletal Committee noted several concerns related to the validity of the measures. Specifically, for #0052: Use of Imaging Studies for Low Back Pain (National Committee for Quality Assurance), the Committee believed that more exclusions for the measure were needed to reflect the "red flag" conditions included in the then-current version of the American College of Radiology guideline for appropriate imaging for low back pain (LBP). For #0514: MRI Lumbar Spine for Low Back Pain (Centers for Medicare and Medicaid Services), the Committee voiced concerns about the ability to identify antecedent conservative therapy through evaluation and management claims, concerns about the 90-day look-back period for prior surgery and trauma, and concerns about the lack of exclusions for patients over 70 years of age.

Upon review of the Standing Committee's recommendation not to re-endorse the measures, the Consensus Standards Approval Committee (CSAC) noted the low frequency of many of the exclusions included in the measures and questioned the need for additional exclusions that likely also would be low-frequency. Ultimately, the CSAC deferred a final endorsement decision, in an effort to give both measure developers time to address the Committee's concerns.

For this current project, developers of measure #0052 revised the specifications to include physical therapy and telehealth visits, shortened the look-back period for the exclusion due to recent trauma,

and excluded those with prolonged use of corticosteroids, HIV, major organ transplant, and recent spinal infection. However, the Committee remained concerned the data element validity testing results, which indicated that as many as ten percent of patients with trauma and five percent of patients with neurologic impairment were not captured using administrative claims data and therefore not excluded from the measure. Although the developers provided updated data from two health plans to demonstrate the impact of the changes in the measure specifications on the measure denominator, exclusions, and performance rate, Committee members noted these data did not explicitly address the adequacy of capture of trauma and neurologic impairment in claims data.

For measure #0514, the Committee expressed confusion about what changes, if any, have been made to the measure since the 2014 evaluation. Although the developers described the various analyses they performed (e.g., quantitative and qualitative evaluation of the look-back periods for several of the measure exclusions), it was still not clear to the Committee if or how the measure has been revised. The Committee expressed concerns with the exclusions and the continued inclusion of "elderly" patients in the measure. The Committee also continued to have concerns with using administrative claims data to identify use of antecedent conservative therapies.

DRAFT REPORT

The *Musculoskeletal: Off-Cycle Measure Review 2017 Draft Report* presents the results of the evaluation of two measures considered under the Consensus Development Process (CDP). During their initial evaluation of the measures, the Committee did not recommend the measures for endorsement. During the post-comment call, the Committee discussed and re-voted on the two measures, but maintained the decision not to recommend the measures for re-endorsement.

	Maintenance	New	Total
Measures under consideration	2	0	2
Measures not recommended for endorsement	2	0	2
Reasons for not recommending	Importance- 0 Scientific Acceptability- 2 Overall- 0 Competing Measure- 0	Importance- N/A Scientific Acceptability- N/A Overall- N/A Competing Measure- N/A	

CSAC ACTION REQUIRED

Pursuant to the CDP, the CSAC may consider endorsement of two candidate consensus standards.

Musculoskeletal Measures Not Recommended (See Appendix A for the Committees votes and rationale)

- 0052: Use of Imaging Studies for Low Back Pain (National Committee for Quality Assurance)
- 0514: MRI Lumbar Spine for Low Back Pain (Centers for Medicare and Medicaid Services)

COMMENTS AND THEIR DISPOSITION

NQF received 10 comments from four member organizations and individuals pertaining 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 CDP Standing Committee Off-Cycle Activities project page.

Comment Themes and Committee Responses

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

During the April 6, 2017 Post-Comment Call, the Standing Committee reviewed all of the submitted comments and developer responses.

Measure Specific Comments

0052: Use of Imaging Studies for Low Back Pain (National Committee for Quality Assurance)

NQF received five post-evaluation comments regarding this measure. (Note: One Musculoskeletal Standing Committee member submitted a comment.) Three of the commenters supported the decision of the Committee not to endorse the measure. Commenters emphasized the importance of limiting unnecessary imaging for low back pain, but expressed concerns over the exclusions and the validity of the measure. Two of the commenters supported of the measure.

Committee Response: NQF staff directed the Committee to re-vote on the validity subcriterion during the post-comment call because of the change to staff preliminary rating for validity from "Insufficient" to "Moderate" after the Committee's initial evaluation. Staff initially noted that only percentage agreement statistics were provided to show the level of agreement between administrative codes and medical records and that the results provided were not calculated using the newly-specified measure. After further consideration of the analysis of the ability to identify the new exclusions in claims only, in medical records only, or in both, staff determined that these data shed light on questions that are addressed in sensitivity/specificity analysis, even though the developer did not provide actual sensitivity/specificity statistics. However, the Committee remained concerned the data element validity testing results, which indicated that as many as ten percent of patients with trauma and five percent of patients with neurologic impairment were not captured using administrative claims data and therefore not excluded from the measure. Although the developers provided updated data from two health plans to demonstrate the impact of the changes in the measure specifications on the measure denominator, exclusions, and performance rate, Committee members noted that these data did not explicitly address the adequacy of capture of trauma and neurologic impairment in claims data. After a lengthy discussion of the initial testing results and updated data, the Committee ultimately agreed that the measure did not pass the validity subcriterion and therefore did not recommend the measure for endorsement.

0514: MRI Lumbar Spine for Low Back Pain (Centers for Medicare and Medicaid Services)

NQF received five post-evaluation comments regarding this measure. (Note: One Musculoskeletal Standing Committee member submitted a comment.) Three of the commenters supported the decision of the Committee not to endorse the measure. Two of commenters supported the measure. Commenters emphasized the importance of limiting unnecessary imaging for low back pain, but expressed concerns over the exclusions and the validity of the measure.

One commenter—the developer of the measure—formally requested a reconsideration of the validity subcriterion. In the request for reconsideration, the developer stated that the "*measure specifications are aligned with the most updated clinical practice guidelines and have strong face validity; additionally, measure testing confirms that threats to validity have been addressed by the exclusion of red-flag conditions*". Because it was unclear to the Committee what changes have been made to the measure since the 2014 review, the developer also clarified that updates to the specifications include the addition of congenital spine/spinal cord malformations, inflammatory and autoimmune disorders, infectious conditions, spinal vascular malformations, spinal cord infarctions, effects from radiation, spinal abnormalities associated with scoliosis, syringohydromyelia, and postoperative fluid collections/soft tissue changes, all of which were added to the measure's list of exclusions.

Committee Response: After a full-discussion of the validity subcriterion, Committee members motioned to reconsider the measure. The Committee again expressed concerns with using administrative claims data to identify use of antecedent conservative therapies and continued to question several of the look-back periods for some of the exclusions (e.g., 90 days for spine surgery and 12 months for cancer). Committee members were also concerned that the specifications do not include certain diagnoses codes to account for several disease states that should be excluded from the denominator (e.g., sciatica and radicular pain, and degenerative conditions). The Committee re-voted on validity and the measure did not pass this subcriterion.

Appendix A-Measure Evaluation Summary Tables

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

0052 Use of Imaging Studies for Low Back Pain

Submission

Description: The percentage of patients with a primary diagnosis of low back pain who did not have an imaging study (plain X-ray, MRI, CT scan) within 28 days of diagnosis.

Numerator Statement: Patients who received an imaging study (x-ray, CT, MRI) within the 28 days following a diagnosis of low back pain.

Denominator Statement: All patients 18 years as of January 1 of the measurement year to 50 years as of December 31 of the measurement year with a claim/encounter for an outpatient, observation, emergency department, physical therapy, or telehealth visit, or osteopathic or chiropractic manipulative treatment, with a principal diagnosis of low back pain during the Intake Period (January 1 – December 3 of the measurement year).

Exclusions: Because the intent of the measure is to assess imaging for patients with a new episode of low back pain, exclude patients with a recent diagnosis of low back pain.

Also, exclude any patient who had a diagnosis for which imaging is clinically appropriate. Any of the following meet criteria:

(1) Cancer
(2) Trauma
(3) Recent IV drug abuse
(4) Neurologic impairment
(5) HIV
(6) Spinal infection
(7) Major organ transplant
(8) Prolonged use of corticosteroids

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Health Plan; Integrated Delivery System

Setting of Care: Clinician Office/Clinic; Emergency Department; Urgent Care - Ambulatory

Type of Measure: Process

Data Source: Claims (Only); This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Provider Organizations via NCQA's online data submission system.

Measure Steward: National Committee for Quality Assurance

STEERING COMMITTEE MEETING [12/12/16]

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-0**; **M-16**; **L-3**; **I-0**

Rationale:

- The developer updated the evidence to include the 2015 American College of Radiology (ACR) Appropriateness Criteria: Low Back Pain. The Committee agreed that this was an appropriate update to

0052 Use of Imaging Studies for Low Back Pain

the evidence and there was no need to re-discuss and re-vote on the evidence sub-criterion.

- Although the developers were unable to provide current performance rates for the measure as specified, the Committee agreed that rates of inappropriate imaging of patients with low back pain continues to be relatively high.
- The developer referenced a recent study from the Department of Veterans Affairs, which found significantly higher rates of MRI in younger adults compared to older adults and significantly lower rates in blacks compared to whites.

2. Scientific Acceptability of Measure Properties: <u>The measure does not meet the Scientific Acceptability criteria</u>

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

2a. Reliability: **H-0**; **M-1**; **L-0**; **I-17** 2b. Validity: **H-0**; **M-0**; **L-5**; **I-12**

<u>Rationale</u>:

- As a response to Committee feedback during the previous evaluation of the measure, the developer revised the specifications to expand the population being measured (i.e., including physical therapy and telehealth visits, shortening the look-back period for the exclusion due to recent trauma, and expanding the list of exclusions to include those with prolonged use of corticosteroids, HIV, major organ transplant, and recent spinal infection).
- Committee members had several questions and concerns about the measure specifications, as follows:
    - Members questioned not requiring pharmacy data to identify those with prolonged use of corticosteroids. The developers acknowledged that use of pharmacy claims could potentially identify additional patients who should be excluded from the measure, but noted that a requirement for the pharmacy benefit would reduce the number of plans that could report on the measure. They stated that, according to their testing information, the rate of prolonged corticosteroid use is quite low, suggesting that lack of pharmacy data does not substantively impact the measure results.
    - Members noted a preference for excluding any patient with a history of spine surgery. The developers noted that the measure can capture anyone with spine surgery in the 6 months prior to the measure index date.
    - Members asked whether patients with radicular symptoms would or would not be excluded from the measure. The developers noted and the Committee agreed that guidelines indicate these patients should not have imaging in the first six weeks. They clarified that these patients would *not* be excluded from the measure via the neurologic impairment exclusion (i.e., they would be included in the measure).
    - One member noted that spinal infection often is identified via imaging. The developers noted that if a patient has a diagnosis within 28 days of the imaging study that indicates spinal infection, that patient is excluded from the measure.
    - Committee members questioned the 28-day post-imaging threshold for exclusions, noting that the timeframe seems somewhat arbitrary and suggesting that it might not be long enough if LBP is treated with conservative management at initial diagnosis. The developers clarified that the 28-day threshold is applied after the imaging study, not after the initial diagnosis.
    - Committee members asked for clarifications about the timeframe for trauma exclusions and about what, specifically, is identified as “trauma.” The developers noted that the value set included as part of the measure specifications include a list of codes used for the trauma exclusions. They also clarified that the timeframe for trauma exclusions is within three months prior to the diagnosis of LBP.

0052 Use of Imaging Studies for Low Back Pain

  - One member noted that the literature indicates that patients with known anatomic spinal anomalies may not require imaging, although this is not covered in the guidelines. The developers noted they have crafted the measure to align with current guidelines.
- Updated score-level testing results based on the revised specifications were not provided. The developer noted that these data would not be available for analysis until mid-2017.
- Although the developers did not provide updated data-element level validity testing, they did provide additional information from their 2002 field testing analysis; these data provided insight on the ability to identify patients in the recently-added exclusion categories (i.e., prolonged steroid use, spinal infection, and immunosuppression). The developer stated that according to the administrative data from 2002, the various red-flag conditions specified as exclusions in the measure occur in 0% to 1.9% of LBP episodes (4.9% overall).
- Committee members expressed concern that a significant number of patients with trauma or neurologic impairment are not being captured using administrative claims data. The developer responded that the testing was performed using 2003-2004 data. They suggested the possibility that claims for trauma and neurologic impairment may have improved since then. They also noted a lack of feedback from health plans that the measure is actually missing a lot of trauma cases.
- In their submission materials, the developer presented their expert panel and commenting processes as an assessment of face validity. As part of the discussion, they provided additional explanation of how they believe these processes fulfill NQF's requirement for face validity.
- The Committee's rating of reliability had to depend, to a large extent, one the data element validity testing information provided by the developer (because the developer was unable to update their score-level reliability testing at this time). Therefore, NQF staff allowed for a vote on validity prior to the vote on reliability. Ultimately, the Committee agreed that there was insufficient information provided for validity and did not recommend the measure for endorsement.

STAFF NOTE: In the staff preliminary analysis and during the measure evaluation webinar, NQF staff noted that only percentage agreement statistics were provided to show the level of agreement between administrative codes and medical records and that the results provided were not calculated using the newly-specified measure. However, after further reflection on the additional information that was provided as part of the exclusion analysis (specifically, data on the ability to identify the new exclusions in claims only, in medical records only, or in both), staff determined that these data shed light on questions that are addressed in sensitivity/specificity analysis, even though the developer did not provide actual sensitivity/specificity statistics. Thus, staff no longer considers the data element validity testing presented by the developer to be insufficient. Because we have reversed our previous guidance, NQF will ask the Committee to reconsider and re-vote on Reliability and Validity during the post-comment call. Also, the developer is seeking updated data to present during the post-comment call.

3. Feasibility: H-12; M-7; 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)

STAFF NOTE: Even though this measure did not pass the Validity subcritieria, staff asked the Committee to use the remaining time on the webinar to discuss and vote on the Feasibility and Usability and Use criteria, given that the developer would be providing additional data during the post-comment call.

Rationale:

- All data elements are in defined fields in electronic claims and generated or collected by and used by

0052 Use of Imaging Studies for Low Back Pain

healthcare personnel during the provision of care. The Committee agreed that the data are readily available and can be captured without undue burden.

4. Usability and Use: H-0; M-18; 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:

- This measure is being used in at least 10 accountability programs, including pay-for-performance programs, accreditation programs, and public reporting.
- One member asked if there might be unintended consequences of the measure if patients who meet red-flag conditions from the guideline are not being excluded from the measure. The developers acknowledged the possibility, but emphasized their process of obtaining feedback from plans.
- The Committee expressed concern about the lack of improvement in performance in the last several years. The developer expressed hope that the Choosing Wisely Campaign would help promote more attention to the issue and drive improvement in performance of the measure.

5. Related and Competing Measures

This measure is competing with:

- #0514: MRI Lumbar Spine for Low Back Pain (CMS)
  - Due to differences in the level of analysis and care settings, the Committee will not be asked to select a best in-class measure.
  - Since the last evaluation, the developers have worked to harmonize the measures, resulting in greater congruence in how low back pain is defined, how cancer exclusions are defined, and in exclusion categories. Areas where the measures are not yet harmonized include the specific imaging modalities captured by the measure and some of the exclusion categories.

Steering Committee Recommendation for Endorsement: **Not Recommended**

6. Public and Member Comment

- NQF received five post-evaluation comments regarding this measure. (Note: One Musculoskeletal Standing Committee member submitted a comment.) Three of the commenters supported the decision of the Committee not to endorse the measure. Two of commenters supported the measure.
- Commenters emphasized the importance of limiting unnecessary imaging for low back pain, but expressed concerns over the exclusions and the validity of the measure.

NQF Post Comment Call

- NQF staff directed the Committee to re-vote on the validity subcriterion during the post-comment call because of the change to staff preliminary rating for validity from “Insufficient” to “Moderate” after the Committee’s initial evaluation. As noted above, staff initially noted that only percentage agreement statistics were provided to show the level of agreement between administrative codes and medical records and that the results provided were not calculated using the newly-specified measure. After further consideration of the analysis of the ability to identify the new exclusions in claims only, in medical records only, or in both, staff determined that these data shed light on questions that are addressed in sensitivity/specificity analysis, even though the developer did not provide actual sensitivity/specificity statistics.
- During the public and member commenting period, the developer provided data from two health plans to demonstrate the impact of the changes in the measure specifications on the measure denominator, exclusions, and performance rate.
- Committee members remained concerned the data element validity testing results, which indicated

0052 Use of Imaging Studies for Low Back Pain
that as many as ten percent of patients with trauma and five percent of patients with neurologic impairment were not captured using administrative claims data and therefore not excluded from the measure. • Although the developers provided updated data, Committee member noted that these data did not explicitly address the adequacy of capture of trauma and neurologic impairment in claims data. • The developer again suggested the possibility that claims for trauma and neurologic impairment may have improved since the field-testing was conducted in 2004. They also noted a lack of feedback from health plans that the measure is actually missing many trauma cases. • After a lengthy discussion of the initial testing results and updated data, the Committee ultimately agreed that the measure did not pass the validity subcriterion and therefore did not recommend the measure for endorsement. **Vote Following Consideration of Public and Member Comments:** **Validity: H-0; M-5; L-4; I-4**
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Appeals

0514 MRI Lumbar Spine for Low Back Pain

Submission

Description: This measure evaluates the percentage of magnetic resonance imaging (MRI) of the lumbar spine studies for low back pain performed in the outpatient setting where conservative therapy was not attempted prior to the MRI. Antecedent conservative therapy may include claim(s) for physical therapy in the 60 days preceding the lumbar spine MRI, claim(s) for chiropractic evaluation and manipulative treatment in the 60 days preceding the lumbar spine MRI, or claim(s) for evaluation and management at least 28 days but no later than 60 days preceding the lumbar spine MRI. The measure is calculated based on a one-year window of Medicare claims data. The measure has been publicly reported, annually, by the measure steward, the Centers for Medicare & Medicaid Services (CMS), since 2010, as a component of its Hospital Outpatient Quality Reporting (HOQR) Program.

Numerator Statement: MRI of the lumbar spine studies with a diagnosis of low back pain (from the denominator) without the patient having claims-based evidence of prior antecedent conservative therapy.

Denominator Statement: The number of MRI of the lumbar spine studies with a diagnosis of low back pain on the imaging claim performed in a hospital outpatient department on Medicare FFS beneficiaries within a 12-month time window.

Exclusions: Indications for measure exclusion include any patients with the following diagnosis code categories:

-Patients with lumbar spine surgery in the 90 days prior to MRI

-Cancer (within twelve months prior to MRI procedure)

-Congenital spine and spinal cord malformations (within five years prior to MRI procedure)

-Inflammatory and autoimmune disorders (within five years prior to MRI procedure)

-Infectious conditions (within one year prior to MRI procedure)

-Spinal vascular malformations and/or the cause of occult subarachnoid hemorrhage (within five years prior to MRI procedure)

-Spinal cord infarction (within one year prior to MRI procedure)

-Neoplastic abnormalities (within five years prior to MRI procedure)

-Treatment fields for radiation therapy (within five years prior to MRI procedure)

-Spinal abnormalities associated with scoliosis (within five years prior to MRI procedure)

-Syringohydromyelia (within five years prior to MRI procedure)

-Postoperative fluid collections and soft tissue changes (within one year prior to MRI procedure)

-Trauma (within 45 days prior to MRI procedure)

-IV drug abuse (within twelve months prior to MRI procedure)

-Neurologic impairment: (within twelve months prior to MRI procedure)

-HIV (within twelve months prior to MRI procedure)

-Unspecified immune deficiencies (within twelve months prior to MRI procedure)

-Intraspinal abscess (an exclusion diagnosis must be in one of the diagnoses fields on the MRI lumbar spine claim)

(Specific CPT codes, ICD-9 codes, and ICD-10 codes for exclusion are included in the value sets for this measure; this detailed list can be found in the Excel workbook provided for criterion S2b.)

Adjustment/Stratification: No risk adjustment or risk stratification

Level of Analysis: Facility; Population : Regional and State

Setting of Care: Emergency Department, Clinician Office/Clinic, Hospital : Hospital, Hospital : Acute Care Facility,

0514 MRI Lumbar Spine for Low Back Pain

Hospital : Critical Care, Imaging Facility, Urgent Care - Ambulatory

Type of Measure: Process

Data Source: Claims (Only)

Measure Steward: Centers for Medicare and Medicaid Services

STEERING COMMITTEE MEETING [1/6/2017]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

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

1a. Evidence: **Previous Evidence Evaluation Accepted**; 1b. Performance Gap: **H-3**; **M-10**; **L-0**; **I-0**

<u>Rationale</u>:

- The developer updated the evidence to include the 2015 American College of Radiology (ACR) Appropriateness Criteria: Low Back Pain. The Committee agreed that this was an appropriate update to the evidence and there was no need to re-discuss and re-vote on the evidence sub-criterion.
- To demonstrate opportunity for improvement, the developer provided an analysis of Medicare fee-for-service (FFS) claims data that indicates variation in the use of inappropriate MRI lumbar spine studies. Performance rates for July 2104 to June 2015 averaged 39.5% and ranged from 14.9% to 64.8% (NOTE: a lower rate is better).
- Committee members noted that the performance gap data actually demonstrated a *decrease* in performance (from 32.5% in 2009 to 39.5% in 2014-2015). The developer indicated that this could be a result of a change in data sources that were used to compute performance scores. The developer also noted that changes in specifications over time make it difficult to interpret changes in performance across time (specifically, expanding the exclusions would decrease the measure denominator, but would not uniformly affect the measure result).
- 2013 data presented by the developer showed that beneficiary age, gender, and race, as well as facility characteristics (i.e., number of beds, urban/rural locality, teaching status) were significantly associated with the rate of inappropriate MRI lumbar spine studies.

2. Scientific Acceptability of Measure Properties: <u>The measure does not meet the Scientific Acceptability criteria</u>

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

2a. Reliability: **H-0**; **M-8**; **L-5**; **I-0** 2b. Validity: **H-0**; **M-3**; **L-9**; **I-1**

<u>Rationale</u>:

- Committee members had several questions and concerns about the measure specifications, as follows:
    - The measure is specified for Medicare Fee-for-Service beneficiaries. However, "elderly individuals" is one of the red-flag conditions in the Appropriate Use guideline, indicating that imaging for the patients presenting with LBP may be appropriate. The developer interpreted the guideline as indicating that "elderly" should not be an independent indicator for imaging; however, some Committee members disagreed with this interpretation.
    - The measure uses evaluation and management (E&M) visits as a proxy for antecedent conservative care (in addition to claims for physical therapy or chiropractic visits). In general, the Committee agreed that the E&M visits are a reasonable proxy for some kinds of antecedent therapy, but questioned whether they would capture other types of antecedent therapy such as telephone encounters. Members noted that some types of antecedent conservative care (e.g., NSAIDs, Tylenol, massage therapy, acupuncture) cannot be captured in claims data.
    - Members questioned several of the look-back periods for some of the exclusions (e.g., 90 days for spine surgery, 12 months for cancer; 5 years for congenital spine and spinal cord malformations). For congenital malformations, the developer clarified that the 5-year look-

0514 MRI Lumbar Spine for Low Back Pain

back was mainly because of lack of access to historical data.

- Committee members expressed concern that specific codes for neurological impairment, specifically those for which the evidence supports appropriate use of MRI, are not adequately captured in this measure. The developer agreed to look into the coding, but also noted that the red flag conditions often occur in tandem, meaning individual patients often are excluded from the measure due to several of the existing measure exclusions. Committee members noted that sciatica radiculopathy, typically does not present with other red-flag conditions.

- The Committee expressed confusion about what changes, if any, have been made to the measure since the 2014 evaluation. Although the developers described the various analyses they performed (e.g., quantitative and qualitative evaluation of the look-back periods for several of the measure exclusions), it is still not clear if or how the measure has been revised. Some of the confusion dates back to the 2014 evaluation, when the developer had actually added several exclusions to the measure that were not apparent in the submission materials considered by the Committee.
- The developers presented updated score-level signal-to-noise reliability testing using 2013 data. Reliability scores from this analysis ranged from 22.4% to 86.6%, with a median reliability score of 44.9%. The median value was well below 0.7, which is often used as a rule-of-thumb minimal acceptable value, and lower than the 53.1% found in previous testing. The developers also provided, a couple of days prior to the evaluation webinar, another set of testing results. This new testing used a split-sample (or "test-retest") approach to compare agreement in performance across hospitals. The intraclass correlation coefficient from this analysis was 0.59, which can be interpreted as moderate agreement (i.e., there is moderate consistency in performance within facilities).
- The developers assessed the face validity of the measure score by surveying an 11-member Technical Expert Panel (TEP). They asked the TEP members to indicate whether the measure captures the most appropriate and prevalent types of antecedent conservative therapy available through claims data (8 of 11 said yes) and to indicate their agreement as to whether the measure helps assess the inappropriate use of MRI lumbar-spine tests (9 of 11 agreed or strongly agreed).
- The developer clarified that the intent of the measure is not to drive measure results to zero, but to decrease the number of orders for MRI on presentation of LBP and to reduce variation between facilities in inappropriate MRIs.
- After much discussion, the Committee agreed that the measure did not pass the validity subcriterion and did not recommend the measure for endorsement.

3. Feasibility: H-X; M-X; L-X; I-X

(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:

-

4. Usability and Use: H-X; M-X; L-X; I-X

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

Rationale:

-

5. Related and Competing Measures

This measure is competing with:

- #0052: Use of Imagine Studies for Low Back Pain (NCQA)
  - Due to differences in the level of analysis and care settings, the Committee will not be asked to select a best in-class measure.

0514 MRI Lumbar Spine for Low Back Pain

- Since the last evaluation, the developers have worked to harmonize the measures, resulting in greater congruence in how low back pain is defined, how cancer exclusions are defined, and in exclusion categories. Areas where the measures are not yet harmonized include the specific imaging modalities captured by the measure and some of the exclusion categories.

Steering Committee Recommendation for Endorsement: **Not Recommended**

6. Public and Member Comment

- NQF received five post-evaluation comments regarding this measure. (Note: One Musculoskeletal Standing Committee member submitted a comment.) Three of the commenters supported the decision of the Committee not to endorse the measure. Two of commenters supported the measure.
- Commenters emphasized the importance of limiting unnecessary imaging for low back pain, but expressed concerns over the exclusions and the validity of the measure.
- One commenter—the developer of the measure—formally requested a reconsideration of the validity subcriterion:

  > "The Centers for Medicare & Medicaid Services (CMS) has requested a reconsideration of the National Quality Forum (NQF) Musculoskeletal Standing Committee's decision not to recommend NQF #0514, MRI Lumbar Spine for Low Back Pain, for continued endorsement. NQF #0514 was originally endorsed by the Outpatient Imaging Efficiency Steering Committee in October 2008; during the January 6, 2017 review webinar, it did not pass the Validity criterion.
  >
  > Based on NQF's Measure Evaluation Criteria and Guidance, we believe that NQF #0514 aligns with the moderate validity recommendation from algorithm #3 (Guidance for Evaluating Validity), as it has received in prior evaluations for endorsement. The measure specifications are aligned with the most updated clinical practice guidelines and have strong face validity; additionally, measure testing confirms that threats to validity have been addressed by the exclusion of red-flag conditions. NQF #0514 also passed the Importance and Reliability criteria during endorsement maintenance review. As one Standing Committee member stated during the review webinar, there will always be exceptions in health care, and, as long as the rate of exceptions is low, performance scores will not be impacted and the measure serves its purpose; we believe that, as currently specified, the measure addresses the broader patterns of care".
- In addition, because it was unclear to the Committee what changes have been made to the measure since the 2014 review, the developer clarified that updates to the specifications include the addition of congenital spine/spinal cord malformations, inflammatory and autoimmune disorders, infectious conditions, spinal vascular malformations, spinal cord infarctions, effects from radiation, spinal abnormalities associated with scoliosis, syringohydromyelia, and postoperative fluid collections/soft tissue changes, all of which were added to the measure's list of exclusions.

<u>NQF Post Comment Call</u>

- On the post-draft report comment call, the Committee reviewed the reconsideration request. The Committee agreed to reconsider the measure for endorsement.
- The Committee again expressed concerns with using administrative claims data to identify use of antecedent conservative therapies, noting that many conservative modalities may not be captured, causing a real risk to the validity of the measure.
- The Committee continued to question several of the look-back periods for some of the exclusions (e.g., 90 days for spine surgery, 12 months for cancer).
- Committee members remained concerned that the specifications do not include certain diagnoses

0514 MRI Lumbar Spine for Low Back Pain
codes to account for several disease states (e.g., sciatica and radicular pain, and degenerative conditions). The developer stated that they have not received feedback from the measure's TEP or external stakeholders that suggests these diagnoses should be excluded from the measure's denominator, however, they welcomed the Committee's feedback and will consider it as they continue to refine the measure during future annual updates. • After a full discussion and review of the request for reconsideration, the Committee ultimately agreed that the measure did not pass the validity subcriterion. Therefore, the measure was not recommended for endorsement. **Vote Following Consideration of Public and Member Comments:** **Validity: H-0; M-3; L-8; I-2**
7. Consensus Standards Approval Committee (CSAC) Vote: Y-X; N-X
8. Appeals

NATIONAL QUALITY FORUM

Musculoskeletal: Off-Cycle Measure Review 2017

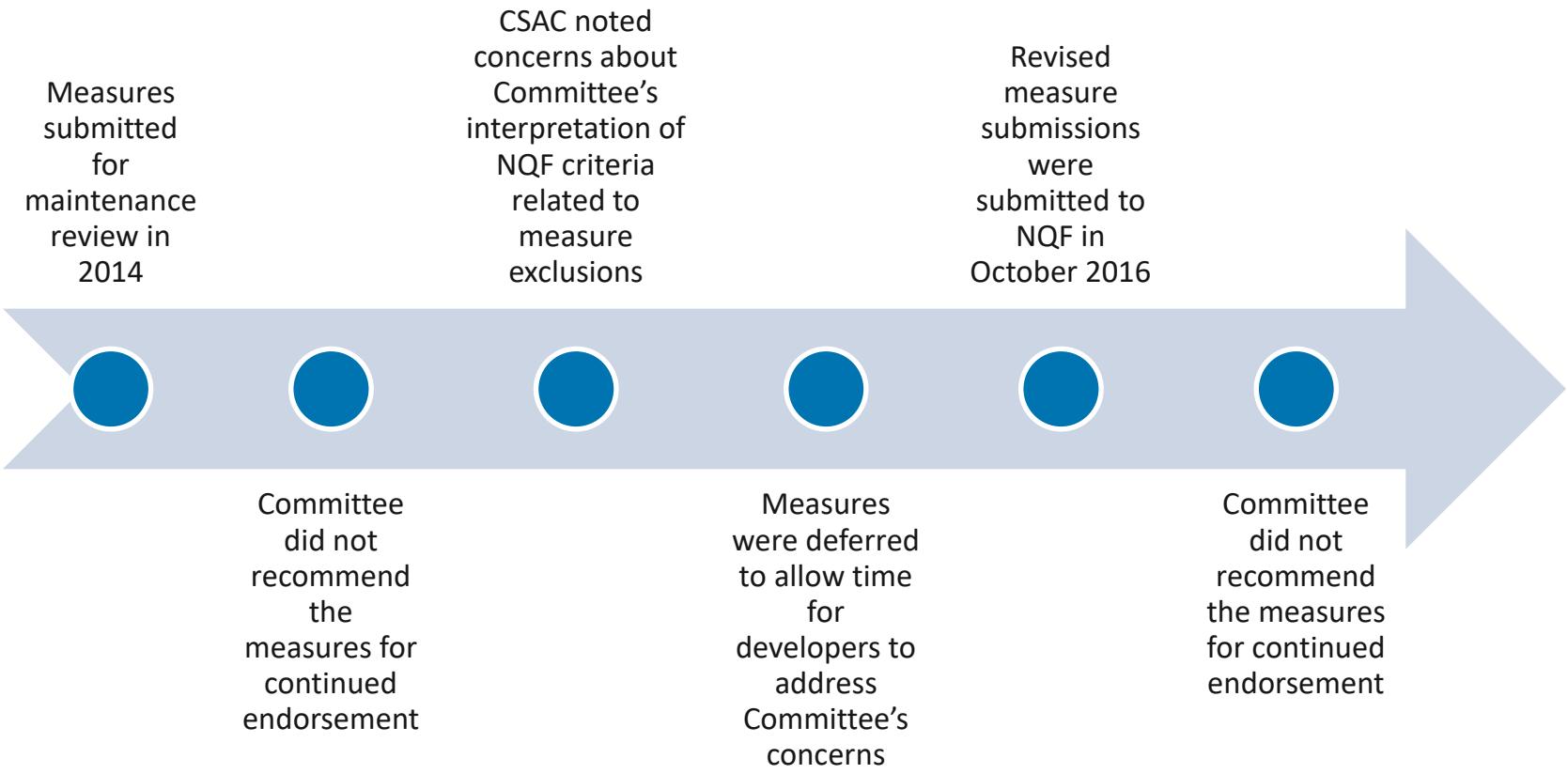
Consensus Standards Approval Committee Review and Recommendations

Roger Chou, MD, FACP (Co-Chair)
Kim Templeton, MD (Co-Chair)

May 1, 2017

Musculoskeletal Off-Cycle Review

0052: Use of Imaging Studies for Low Back Pain
0514: MRI Lumbar Spine for Low Back Pain



Musculoskeletal: Off-Cycle Measure Review 2017

	Maintenance	New	Total
Measures under consideration	2	0	2
Measures not recommended for endorsement	2	0	2
Reasons for not recommending	Importance – 0 Scientific Acceptability –2 Overall – 0 Competing Measure – N/A	Importance – N/A Scientific Acceptability – N/A Overall – N/A Competing Measure – N/A	

Overarching Issues

- Measure Exclusions
- Application of Complex Guidelines to Claims Data
- Expectation of Substantial Changes to Measure Specifications

Comments Received

Musculoskeletal: Off-Cycle Measure Review 2017 Comments Received

- Ten comments received from four NQF member organizations
  - Five comments received on each measure under review
    - » Three commenters supported the decision of the Committee not to endorse the measures
    - » Two commenters supported the measures
    - » One commenter—the developer of measure 0514: *MRI Lumbar Spine for Low Back Pain,* formally requested a reconsideration of the measure due to inappropriate application of the evaluation criteria

Committee Recommendations following Member and Public Comment:

- The Committee upheld its recommendation to not recommend the two measures:
  - 0052: *Use of Imaging Studies for Low Back Pain*
  - 0514: *MRI Lumbar Spine for Low Back Pain*

Musculoskeletal Off-Cycle Measure Review 2017 Project Timeline and Next Steps

Appeals*	May 3 – June 2
Final Report	July 28

*An appeal may be filed on endorsed measures only

Questions?