NATIONAL QUALITY FORUM
Driving measurable health improvements together

Meeting Summary

Cost and Efficiency Standing Committee – Measure Evaluation Web Meeting

The National Quality Forum (NQF) convened the Cost and Efficiency Standing Committee for a web meeting on July 12, 2022, to evaluate three new measures for the spring 2022 cycle.

Welcome, Review of Meeting Objectives, Introductions, and Overview of Evaluation and Voting Process

LeeAnn White, NQF director, welcomed the Standing Committee and participants to the web meeting. NQF staff reviewed the meeting objectives. The Standing Committee members each introduced themselves and disclosed any conflicts of interest. There were no Standing Committee members recused from the measures under review. Additionally, Isaac Sakyi, NQF manager, reviewed the Consensus Development Process (CDP) and the measure evaluation criteria.

During the meeting, the quorum required for live voting was not achieved (10 Standing Committee members). Therefore, the Standing Committee discussed all relevant criteria and voted after the meeting using an online voting tool. The Standing Committee received a recording of the meeting and a link to submit online votes. Voting results are provided below.

Measure Evaluation

During the meeting, the Cost and Efficiency Standing Committee evaluated three new measures for endorsement consideration. A more detailed summary of the Standing Committee's deliberations will be compiled and provided in the draft technical report.

A measure is recommended for endorsement by the Standing Committee when greater than 60 percent of eligible voting members select a passing vote option (Pass, High and Moderate, Yes) on all must-pass criteria and overall suitability for endorsement. A measure is not recommended for endorsement when less than 40 percent of voting members select a passing vote option on any must-pass criterion or overall suitability for endorsement. If a measure does not pass a must-pass criterion, voting during the measure evaluation meeting will cease. The Standing Committee will not re-vote on the measure(s) during the post-comment meeting unless the Standing Committee decides to reconsider the measure(s) based on submitted comments or a formal reconsideration request from the developer. The Standing Committee has not reached consensus on the measure if between 40 and 60 percent of eligible voting members select a passing vote option on any must-pass criterion or overall suitability for endorsement. The Standing Committee will re-vote on criteria that did not reach consensus and potentially on overall suitability for endorsement during the post-comment web meeting.

Voting Legend:

- *Evidence (Outcome Measures) and Use:* Pass/No Pass
- *Accepting Scientific Methods Panel (SMP) Rating and Overall Suitability for Endorsement*: Yes/No
- *All Other Criterion:* H – High; M – Medium; L – Low; I – Insufficient; NA – Not Applicable
- *Maintenance Criteria for Which Standing Committee Decided Additional Discussion/Vote Was Not Needed (Evidence, Reliability, Validity only):* Accepted Previous Evaluation

NQF #3623 Elective Primary Hip Arthroplasty Measure (Centers for Medicare & Medicaid Services [CMS]/Acumen, LLC)

Description: The Elective Primary Hip Arthroplasty episode-based cost measure evaluates a clinician's risk-adjusted cost to Medicare for patients who receive an elective primary hip arthroplasty during the performance period. The measure score is a clinician's risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care during each episode from the 30 days prior to the clinical event that opens or "triggers" the episode, through 90 days after the trigger. Patient populations eligible for the Elective Primary Hip Arthroplasty measure include Medicare beneficiaries enrolled in Medicare Parts A and B; **Level of Analysis:** Clinician: Individual, Clinician: Group/Practice; **Setting of Care:** Ambulatory Care: Clinician Office, Other, Inpatient/Hospital, Ambulatory Care: Clinic/Urgent Care; **Type of Measure**: Cost and Resource use; **Data Source**: Claims

Measure Steward/Developer Representatives at the Meeting

- Sam Bounds, Acumen, LLC.
- Rose Do, Acumen, LLC.
- Joyce Lam, Acumen, LLC.
- Heather Litvinoff, Acumen, LLC.
- Sriniketh Nagavarapu, Acumen, LLC.

Standing Committee Votes

- **High Impact and Opportunity for Improvement:** Total votes-10; H-2; M-8; L-0; I-0 (10/10-100%, Pass)
- **Reliability**: Total votes-10; H-4; M-6; L-0; I-0 (10/10- 100%, Pass)
    - This measure was deemed as complex and was evaluated by the NQF Scientific Methods Panel (SMP).
    - The NQF SMP's rating for Reliability: High (Total votes-8; H-7; M-1; L-0; I-0)
    - Since voting was conducted offline using a web-based voting tool, the Standing Committee provided their own vote for reliability rather than be asked to uphold the SMP's rating.
- **Validity**: Total votes-10; H-0; M-8; L-1; I-1 (8/10- 80%, Pass)
    - This measure was deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF SMP's rating for Validity: Moderate (Total votes-7; H-0; M-5; L-2; I-0)
    - Since voting was conducted offline using a web-based voting tool, the Standing Committee provided their own vote for validity rather than be asked to uphold the SMP's rating.
- **Feasibility**: Total votes-10; H-7; M-3; L-0; I-0 (10/10- 100%, Pass)
- **Use**: Total votes-10; Pass-9, No Pass-1 (9/10- 90%, Pass)
- **Usability**: Total votes-10; H-1; M-6; L-3; I-0 (7/10- 70%, Pass)
- **Standing Committee Recommendation for Endorsement:** Total votes-10; Y-8; N-2 (8/10-80%, Pass)

The Standing Committee recommended the measure for initial endorsement. This group/practice- and individual clinician-level measure was newly submitted for endorsement. This measure is publicly reported in the Quality Payment Program (QPP) Merit-based Incentive Payment System (MIPS).

The Standing Committee reviewed data provided by the developer that demonstrated a high prevalence of total hip arthroplasties representing 0.8 percent for the general population and increasing with age to 1.5 percent at 60 years and 5.9 percent by 90 years of age. During the discussion on opportunities for improvement, the Standing Committee noted that the performance gap data indicated a mean score of 1.03 (standard deviation [SD] of 0.12, interquartile range [IQR] of 0.15) at the clinician group level and a

mean score of 1.00 (SD of 0.12, IQR 0.15) at the individual clinician-level. The Standing Committee agreed that the 0.15 range translated into significant overall cost savings for Medicare if performance on this measure moved from the 75th to the 25th percentile of cost. The Standing Committee cautioned that while a performance gap in spending was present, it was difficult to ascertain actions that clinicians can take to impact this variation and how it relates to overall patient care quality. The Standing Committee ultimately passed the measure on high impact and improvement opportunities.

The Standing Committee noted that the SMP previously reviewed this measure in spring 2021 and it passed with a high rating on reliability and a moderate rating on validity. The Standing Committee agreed with the SMP's evaluation that the developer's signal-to-noise (SNR) and split-sample reliability testing were sufficient and that the testing results indicated a robust measure of score reliability. The Standing Committee noted that the developer conducted empirical and face validity testing at the accountable entity level. During the discussion on validity, the Standing Committee raised several concerns, specifically with the small size of the developer's initial technical expert panel (TEP), the correlation of this measure with a similar NQF-endorsed resource measure instead of a quality measure, and the merits of the attribution and shared accountability for measure performance (i.e., primary and assisting surgeon). The developer explained that the subsequent TEP was more significant (n=29 members) and included experts in musculoskeletal disease management with affiliations in 26 organizations and specialty societies. Regarding the attribution approach, the developer noted that the primary and assisting surgeon are both attributed as they have joint responsibility for the cost measure. Addressing the Standing Committee's concern with the quality and cost correlation, the developer noted that in addition to the correlation analysis performed with NQF #2158 *Medicare Spending Per Beneficiary (MSPB)- Hospital Measure*, they performed correlation analysis with NQF #3495 *Hospital-Wide 30-Day, All-Cause, Unplanned Readmission Rate (HWR) for the Merit-Based Incentive Payment System (MIPS) Eligible Clinician Groups*. During the discussion, the developer reported a Pearson correlation of 0.27 amongst providers with lower costs and complication rates which they considered a medium correlation.

The Standing Committee also expressed concern with the lack of social risk adjustment in the risk model. The developer explained that the measurement results were stratified by dual eligibility status and found that the risk-adjusted cost for both dual-eligible and non-dual eligible episodes increases among providers with higher dual-eligible populations (i.e., providers with higher dual-eligible beneficiaries may perform worse). The developer expressed concern that risk adjusting for dual status could unintentionally remove some of the difference in performance due to provider-level effect versus individual-level effect. Ultimately, the Standing Committee accepted the developer's responses to the concerns raised, agreed with the SMP, and passed the measure on reliability and validity.

The Standing Committee agreed that the measure is feasible, and the data elements required for the measure are readily available and could be captured without undue burden. The Standing Committee acknowledged that this is a new measure, and that the developer did not provide any improvement data. The Standing Committee questioned how clinicians could improve the quality of care while reducing cost when healthcare settings and services are determined by healthcare systems where physicians are employed. The developer explained that clinicians receive field reports containing cost performance categories that can be further broken down into specific services and settings to identify areas of improvement. The Standing Committee accepted the developer's response and passed the measure on feasibility, use, usability, and overall suitability for endorsement.

There were no related and competing measures identified for this measure.

NQF #3625 Non-Emergent Coronary Artery Bypass Graft (CABG) Measure (CMS/Acumen, LLC)

Description: The Non-Emergent CABG episode-based cost measure evaluates a clinician's risk-adjusted cost to Medicare for patients who undergo a CABG procedure during the performance period. The

measure score is the clinician's risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care during each episode from 30 days prior to the clinical event that opens, or "triggers," the episode through 90 days after the trigger. Patient populations eligible for the Non-Emergent CABG measure include Medicare beneficiaries enrolled in Medicare Parts A and B; **Level of Analysis:** Clinician: Individual, Clinician: Group/Practice; **Setting of Care:** Inpatient/Hospital; **Type of Measure**: Cost and Resource use; **Data Source**: Claims

Measure Steward/Developer Representatives at the Meeting

- Sam Bounds, Acumen, LLC.
- Rose Do, Acumen, LLC.
- Joyce Lam, Acumen, LLC.
- Heather Litvinoff, Acumen, LLC.
- Sriniketh Nagavarapu, Acumen, LLC.

Standing Committee Votes

- **High Impact and Opportunity for Improvement:** Total votes-10; H- 1; M-8; L-1; I-0 (9/10-90%, Pass)
- **Reliability**: Total votes- 10; H-1; M-8; L-1; I-0 (9/10- 90%, Pass)
    - This measure was deemed as complex and was evaluated by the NQF Scientific Methods Panel (SMP).
    - The NQF SMP's rating for Reliability: High/Moderate (Total Votes- 8; H-4; M-4; L-0; I-0)
    - Since voting was conducted offline using a web-based voting tool, the Standing Committee provided their own vote for reliability rather than be asked to uphold the SMP's rating.
- **Validity**: Total votes- 10; H-0; M-7; L-2; I-1 (7/10- 70%, Pass)
    - This measure was deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF SMP's rating for Validity: Moderate (Total votes- 8; H-0; M-5; L-3; I-0)
    - Since voting was conducted offline using a web-based voting tool, the Standing Committee provided their own vote for validity rather than be asked to uphold the SMP's rating.
- **Feasibility**: Total votes-10; H-9; M-1; L-0; I-0 (10/10- 90%, Pass)
- **Use**: Total votes-10; Pass-9; No Pass-1 (9/10- 90%, Pass)
- **Usability**: Total votes- 10; H-0; M-7; L-3; I-0 (7/10- 70%, Pass)
- **Standing Committee Recommendation for Endorsement:** Total votes-10; Y-9; N-1 (9/10- 90%, Pass)

The Standing Committee recommended the measure for initial endorsement. This group/practice- and individual clinician-level measure was newly submitted for endorsement. This measure is publicly reported in the QPP MIPS program.

While the Standing Committee acknowledged a high prevalence of non-emergent CABG surgeries among Medicare beneficiaries reflecting substantial Medicare expenditures, they noted an overall downward trajectory and steady decline of CABG cases and mortality. The developer explained that with the advancements in interventional cardiology, the number of CABG procedures would continue to decrease. The Standing Committee did caution that while a performance gap in spending was present, it is difficult to ascertain the actions clinicians can take to impact this variation and how it relates to overall patient care quality. One Standing Committee member noted that the measure aims to reduce the number of avoidable readmissions and appropriate post-acute care and questioned the rationale for making this measure a cost measure instead of a quality measure. The developer explained that the MIPS cost performance category requires measures based on care episode groups. The developer noted

that they selected the CABG episode because it is a high-frequency, high-cost care area. The Standing Committee accepted the developer's rationale and passed the measure on high impact and opportunity for improvement.

The Standing Committee noted that the measure was previously reviewed by the SMP in spring 2021 and passed with a moderate rating on reliability and validity. While the Standing Committee agreed that the reliability testing was robust, one Standing Committee member requested clarification as to why the developer selected the 10-episode case minimum. The developer explained that there was careful consideration to both coverage and reliability when determining case minimum to ensure that smaller providers with lower case volumes are assessed. The Standing Committee agreed with the SMP the reliability testing was appropriate, and the testing results indicated moderate measure score reliability.

The Standing Committee reviewed the validity testing that the developer conducted at the performance measure score level. While the Standing Committee agreed that the validity testing was robust, they raised concerns about the high number of exclusions. The developer explained that the exclusion logic is designed to capture only non-emergent CABG procedures and the exclusions selected to ensure that the measure is not accidentally picking up emergent procedures. The Standing Committee accepted the developer's rationale, agreed that the validity testing was sufficient, and passed the measure on validity.

The Standing Committee agreed that the measure is feasible, and the data elements required for the measure are readily available and could be captured without undue burden. While the Standing Committee acknowledged that this is a new measure and that the developer did not provide any improvement data, they raised concerns about how the measure's performance results can be used to improve care further. Specifically, the Standing Committee questioned how the developer plans to differentiate between natural variation and areas of actual improvement in care. The developer will continue to monitor the impact of the measure and noted that they expect that there will be an early reduction in cost and then a gradual flattening out and convergence across providers. The developer responded that they expect that there will be an early reduction in cost and then a gradual flattening out and convergence across providers. The Standing Committee passed the measure on feasibility and use.

During the discussion of unintended consequences, one Standing Committee member noted that opportunities for significant cost savings might be excluded when outlier cases are eliminated from the data as this may be where the actual waste and inefficiencies reside. The developer clarified that one percent of episodes at both ends of the distribution are excluded where the risk-adjustment model cannot predict cost accurately. The Standing Committee appreciated the developer's response and ultimately passed the measure on usability and overall suitability for endorsement.

There were no related and competing measures identified for this measure.

NQF #3626 Lumbar Spine Fusion for Degenerative Disease, 1-3 Levels Measure (CMS/Acumen, LLC)

Description: The Lumbar Spine Fusion for Degenerative Disease, 1-3 Levels episode-based cost measure evaluates a clinician's risk-adjusted cost to Medicare for patients who undergo surgery for lumbar spine fusion during the performance period. The measure score is the clinician's risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician. This procedural measure includes costs of services that are clinically related to the attributed clinician's role in managing care during each episode from 30 days prior to the clinical event that opens, or "triggers," the episode through 90 days after the trigger. Patient populations eligible for Lumbar Spine Fusion for Degenerative Disease, 1-3 Levels measure include Medicare beneficiaries enrolled in Medicare Parts A and B; **Level of Analysis:** Clinician: Group/Practice, Clinician: Individual; **Setting of Care:** Inpatient/Hospital, Other, Ambulatory Care: Clinic/Urgent Care; **Type of Measure**: Cost and Resource use; **Data Source**: Claims

Measure Steward/Developer Representatives at the Meeting

- Sam Bounds, Acumen, LLC.
- Rose Do, Acumen, LLC.
- Joyce Lam, Acumen, LLC.
- Heather Litvinoff, Acumen, LLC.
- Sriniketh Nagavarapu, Acumen, LLC.

Standing Committee Votes

- **High Impact and Opportunity for Improvement:** Total votes-10; H-4; M-5; L-1; I-0 (9/10- 90%, Pass)
- **Reliability**: Total votes- 10; H-3; M-7; L-0; I-0 (10/10- 100%, Pass)
    - This measure was deemed as complex and was evaluated by the NQF Scientific Methods Panel (SMP).
    - The NQF SMP's rating for Reliability: High/Moderate (Total votes- 8; H-4; M-4; L-0; I-0)
    - Since voting was conducted offline using a web-based voting tool, the Standing Committee provided their own vote for reliability rather than be asked to uphold the SMP's rating.
- **Validity**: Total votes- 10; H-0; M-9; L-0; I-1 (9/10- 90%, Pass)
    - This measure was deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF SMP's rating for Validity: Moderate (Total votes- 8; H-0; M-6; L-2; I-0)
    - Since voting was conducted offline using a web-based voting tool, the Standing Committee provided their own vote for validity rather than be asked to uphold the SMP's rating.
- **Feasibility**: Total votes- 10; H-8; M-2; L-0; I-0 (10/10- 100%, Pass)
- **Use**: Total votes- 10; Pass-10, No Pass-0 (10/10- 100%, Pass)
- **Usability:** Total votes- 10; H-1; M-7; L-1; I-1 (8/10- 80%, Pass)
- **Standing Committee Recommendation for Endorsement:** Total votes- 10; Y-8; N- 2 (8/10- 80%, Pass)

The Standing Committee recommended the measure for initial endorsement. This group/practice- and individual clinician-level measure was newly submitted for endorsement. This measure is reported publicly in the QPP MIPS program.

The Standing Committee reviewed data demonstrating a high prevalence of degenerative lumber conditions affecting more than six million Medicare patients and a total admission expenditure for lumbar spine fusion surgeries exceeding $3.6 billion in 2013. During the discussion on opportunities for improvement, a Standing Committee member questioned what services tend to drive cost-per-case variability. The developer explained that acute readmissions and post-acute care have the most influence on cost. The Standing Committee agreed that this measure captures an area of high impact and resource use that warrants a national performance measure and passed the measure on both criteria.

The Standing Committee noted that the SMP reviewed this measure in spring 2021 and passed with a moderate rating on reliability and validity. The Standing Committee agreed with the SMP evaluation that the developer's SNR and split-sample reliability testing were sufficient and that the testing results indicated moderate measure score reliability. While the Standing Committee agreed that the validity results were robust, a Standing Committee member requested further clarification on how the developer applied the model across three subgroups (i.e., one subgroup for the three distinct levels of procedures). The developer explained that they stratified all episodes into three mutually exclusive subgroups and applied the risk adjustment model separately within each of the three subgroups. The

developer further explained that the three subgroup scores are rolled up at the provider level to calculate the overall measure score. One Standing Committee member noted that base and race data are challenging to parse out from the current risk model, which combines three components (base, dual eligibility status, and race). The Standing Committee member suggested that the developer consider a risk model that only provides a base population plus race. The Standing Committee passed the measure on validity.

The Standing Committee agreed that the measure is feasible, and the data elements required for the measure are readily available and could be captured without undue burden. During the discussion of usability, the Standing Committee raised concern about the potential for undertreatment and the unintended consequences of pain management and opioid prescribing among patients undergoing lumbar spine procedures. The developer explained that the cost drivers are related to adverse outcomes; undertreatment typically results in costly adverse events that the measure will capture within the 90-day postoperative period. The developer further noted that drugs are included in the service assignment and highlighted the importance of opioid use quality measures, which look specifically at prescribing practices and use. The Standing Committee accepted the developer's response and passed the measure on feasibility, use, usability, and overall suitability for endorsement.

There were no related and competing measures identified for this measure.

Public Comment

Ms. White opened the lines for NQF member and public comments. No public or NQF member comments were provided during the measure evaluation meeting.

Next Steps

Tristan Wind, NQF analyst, provided an overview of next steps. NQF will post the draft technical report containing the Standing Committee's discussion and recommendations on August 15, 2022, for public comment for 30 calendar days. The continuous public commenting period with member support will close on September 13, 2022. NQF will reconvene the Standing Committee for the post-comment web meeting in the fall of 2022.