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 10, 2020 to evaluate six new measures.

Welcome, Introductions, and Review of Meeting Objectives

NQF welcomed the Standing Committee and participants to the web meeting. NQF staff reviewed the meeting objectives. Committee members each introduced themselves and disclosed any conflicts of interests. Standing Committee members, Jack Needleman, PhD, and Bijan Borah, MSc, PhD, were recused from voting on scientific acceptability on all six measures due to their involvement on the NQF Scientific Methods Panel (SMP).

During the meeting, some Committee members were unable to attend the entire meeting. There were early departures and late arrivals due to preexisting conflicts, including those related to COVID-19. Quorum was met for the first part of the meeting where the committee discussed all the measure evaluation criteria for the first measure, NQF #3561, and importance to measure and report and reliability criteria for the second measure, NQF #3562. The vote totals reflect the members present and eligible for each vote. Quorum required for voting was not achieved for measure discussion thereafter. Therefore, the Committee discussed all relevant criteria and voted after the meeting using an online voting tool for the rest of the measure criteria for NQF #3562 and the remaining four measures.

Topic Area Introduction and Overview of Evaluation Process

NQF staff provided an overview of the topic area and the current NQF portfolio of endorsed measures. There are currently 10 endorsed measures in the Cost and Efficiency portfolio. Additionally, NQF reviewed the Consensus Development Process (CDP) and the measure evaluation criteria.

Measure Evaluation

During the meeting, the Cost and Efficiency Standing Committee evaluated six new measures for endorsement consideration. A summary of the Committee deliberations will be compiled and provided in the draft technical report. NQF will post the draft technical report on August 14, 2020 for public comment on the NQF website. It will be posted for 30 calendar days.

Rating Scale: H – High; M – Medium; L – Low; I – Insufficient; NA – Not Applicable

3561 Medicare Spending Per Beneficiary – Post Acute Care Measure for Inpatient Rehabilitation Facilities [Centers for Medicare and Medicaid/Acumen, LLC]

Measure Steward/Developer Representatives at the Meeting

Sri Nagavarapu, Acumen LLC
Alan Levitt, Centers for Medicare and Medicaid Services (CMS)
Ronique Evans, Centers for Medicare and Medicaid Services (CMS)

Standing Committee Votes

- Importance to Measure and Report H-3**;** M-14**;** L-1**;** I-0
- Reliability Y-14; N-1
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF Scientific Methods Panel's rating for Reliability: Moderate (H-3; M-4; L-0; I-0)
    - The Committee accepted the NQF Scientific Methods Panel's rating.
- Validity: H-0**;** M-5**;** L-10**;** I-0
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF Scientific Methods Panel's ratings for Reliability: Moderate (H-1; M-6; L-1; I-0)
    - The Committee did not accept the NQF Scientific Methods Panel's Moderate rating: Yes-8; No-7
- Feasibility: Vote not taken
- Use: Vote not taken
- Usability: Vote not taken

Standing Committee Recommendation for Endorsement: Vote not taken

The Standing Committee did not vote on the recommendation for endorsement at the meeting because the Committee did not pass on validity—a must-pass criterion.

Several Committee members noted that the measure was developed as a part of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act) and is a legislative requirement. The Committee reviewed publicly reported measure score data provided by the developer for all U.S. providers under Medicare's inpatient rehabilitation facilities (IRF) Prospective Payment System (PPS) with 20 or more eligible episodes in the reporting period of 2016-2017. The Committee acknowledged that the developer demonstrated significant variability in resource use across IRFs and agreed that this measure addresses a high resource use aspect of healthcare and that there is an opportunity for improvement. The Committee passed the measure on the importance to measure and report criterion.

The Committee noted that this measure has been evaluated by the SMP and was given passing ratings for both reliability and validity. It noted that the developers demonstrated reliability using signal-to-noise analysis through the Adams' method. The Committee reviewed the reliability results and agreed with the SMP, ultimately passing the measure on reliability. It raised several concerns regarding the validity of the measure. The Committee questioned the use of 30 days as the appropriate length of time that IRFs can influence downstream care decisions. The Committee highlighted the need to empirically evaluate and validate if 30-days post discharge is an appropriate length of time to capture complications that can be attributed to IRF care. The Committee raised concern regarding the approach to truncation/winsorization of low-and-high cost episodes and questioned the approach to how death is handled within the episode window. It also questioned how well the model predicts downstream cost and raised concern regarding the lack of adjustment for social risk factors. The developer noted that

testing demonstrated significance of the social factors but inconsistent direction. It also demonstrated limited impact of social risk factor effects under the current risk adjustment model. The Committee highlighted that accounting for social risk factors associated with the outcome of interest that are outside provider's control reduces bias in measurement.

Lastly, the Committee raised concern that the expected costs were not aligned with how patient risk is accounted for in IRF payment programs. The developer noted on the call that there is more alignment in the expected episode cost and the payment program for long-term care hospitals (NQF #3562) and less for IRF and the other post-acute care (PAC) measures. The Committee did not vote to pass the measure on the validity criterion, noting additional concerns beyond the SMP review related to the lack of social adjustment in the risk adjustment model, alignment of patient risk between expected costs and IRF payment programs, and additional threats to validity as stated above.

3562 Medicare Spending Per Beneficiary – Post Acute Care Measure for Long-Term Care Hospitals [Centers for Medicare and Medicaid/Acumen, LLC]

Measure Steward/Developer Representatives at the Meeting

Sri Nagavarapu, Acumen LLC
Alan Levitt, Centers for Medicare and Medicaid Services (CMS)
Ronique Evans, Centers for Medicare and Medicaid Services (CMS)

Standing Committee Votes

- Importance to Measure and Report: H-2; M-14; L-0; I-0
- Reliability**:** Y-15; N-0
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF Scientific Methods Panel's rating for Reliability: Moderate (H-5; M-2; L-0; I-0)
    - The Committee accepted the NQF Scientific Methods Panel's rating.
- Validity: H-0**;** M-10**;** L-6**;** I-0
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF Scientific Methods Panel's ratings for Reliability: Moderate (H-2; M-3; L-2; I-0)
- Feasibility: H-7; M-11; L-0; I-0
- Use: Pass-18; No Pass-0
- Usability: H-1; M-11; L-5; I-1

Standing Committee Recommendation for Endorsement: Yes-10; No-6

The Standing Committee recommended the measure for initial endorsement.

Several Committee members noted that the measure was developed as a part of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act) and is a legislative requirement. The Committee acknowledged that the developer demonstrated significant variability in resource use in long-term care hospitals with opportunities for improvement. It did note that there is geographic

variability in the availability of long-term care hospital (LTCH) settings across the country. The Committee generally agreed that this measure addresses a high resource use aspect of healthcare. They also acknowledged that the developer demonstrated variation in post-acute care spending to warrant measurement. The Committee passed the measure on importance to measure and report criterion.

The Committee noted that this measure has been evaluated by the SMP and was given passing ratings for both reliability and validity. The Committee noted that the developers demonstrated reliability using signal-to-noise analysis through Adam's method. The Committee reviewed the reliability results and agreed with the SMP that this measure was reliable and voted to pass the measure on reliability. The Committee expressed some concerns related to the validity of the measure. It questioned the use of 30 days as the appropriate length of time that LTCHs can influence downstream care decisions. The Committee highlighted the need to empirically evaluate and validate if 30-days post-discharge period is the appropriate length of time to capture complications that can be attributed to LTCHs. It raised concerns that the calculation of expected cost is not aligned with LTCH payment programs. The developer noted on the call that there is more alignment in the expected episode cost and the payment program for LTCH and less for IRF and the other PAC measures. With respect to risk adjustment, the Committee acknowledged that the adjusted r-squared for this LTCH measure is higher than the other PAC measures (r-squared value of 0.4894). Additionally, the Committee highlighted the importance of adjusting for social risk factors to reduce bias in measurement. Ultimately, the Committee passed this measure on validity.

The Committee regarded the measure as feasible with no concerns. The Committee passed the measure on use and usability, acknowledging that this is a new measure and is publicly reported as part of the Centers of Medicare & Medicaid Services' LTCH Quality Reporting Program. The Committee will discuss related and competing measures during its post-comment web meeting on October 1, 2020. The Committee ultimately passed this measure on overall suitability for endorsement.

3563 Medicare Spending Per Beneficiary – Post Acute Care Measure for Skilled-Nursing Facilities [Centers for Medicare and Medicaid/Acumen, LLC]

Measure Steward/Developer Representatives at the Meeting

Sri Nagavarapu, Acumen LLC
Alan Levitt, Centers for Medicare and Medicaid Services (CMS)
Ronique Evans, Centers for Medicare and Medicaid Services (CMS)

Standing Committee Votes

- Importance to Measure and Report: H-4; M-13; L-0; I-0
- Reliability: H-3; M-11; L-1; I-0
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF Scientific Methods Panel's ratings for Reliability: High/Moderate/Low (H-5; M-3; L-0; I-0)
- Validity: H-0; M-8; L-7; I-0
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF Scientific Methods Panel's ratings for Reliability: High/Moderate/Low (H-2; M-

4; L-1; I-1)

- Feasibility: H-5; M-11; L-1; I-0
- Use: Pass-16; No Pass-1
- Usability: H-0; M-11; L-5; I-1

Standing Committee Recommendation for Endorsement: Vote not taken

The Standing Committee did not vote on the recommendation for endorsement at the meeting because it did not reach consensus on validity—must-pass criterion. The Committee will revote on the measure on the post-comment web meeting on October 1, 2020.

The Committee acknowledged that the measure was developed to address the resource use aspect of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act). The Committee reviewed a range of data demonstrating high impact through differences in PAC payments across skilled nursing facilities (SNFs). It agreed that the range of performance demonstrated an opportunity for improvement in reducing the variability in spending. Overall, the Committee passed the measure on the importance of the measure and report criterion.

The Committee noted that this measure has been evaluated by the SMP, which gave passing ratings for both reliability and validity. For reliability, the Committee reviewed the signal-to-noise analysis and split sample reliability testing conducted by the developer. For reliability, the Committee did not raise any major concerns and passed the measure on this criterion. In terms of validity, the Committee reviewed the exclusion of clinically unrelated services provided by the developer and commented on the importance of excluding certain downstream costs not associated with SNF care. The Committee also had concerns that the calculation of expected cost was not aligned with SNF payment programs. The developer noted on the call that there is more alignment in the expected episode cost and the payment program for LTCH and less for the SNF and the other PAC measures. The Committee also questioned how well the model predicts downstream costs (r-squared of 0.11) and raised concerns regarding the lack of including social factors. The developer indicated that though each of the social factors tested was statistically significant in the model, they did not improve the model fit and the adjusted r-squared values increased by less than 0.01. The Committee stressed that inclusion of risk factors should be about minimizing bias and may not always improve model fit. Due to these concerns, the Committee did not reach consensus on validity.

The Committee passed the measure on feasibility, acknowledging that the measure data are routinely collected and that this measure poses no additional data collection burden on providers. The Committee acknowledged that this is a new measure and is publicly reported as part of the Centers of Medicare & Medicaid Services' SNF Quality Reporting Program. There was some concern that there is a lack of clarity on whether providers have enough information to target improvement, as there was no indication of areas of high or low spending by provider associated with SNF. However, the Committee ultimately passed the measure on use and usability.

3564 Medicare Spending Per Beneficiary – Post Acute Care Measure for Home Health Agencies [Centers for Medicare and Medicaid/Abt Associates]

Measure Steward/Developer Representatives at the Meeting

Morris Hamilton, Abt Associates
Alrick Edwards, Abt Associates
Alan Levitt, Centers for Medicare and Medicaid Services (CMS)

Ronique Evans, Centers for Medicare and Medicaid Services (CMS)

Standing Committee Votes

- Importance to Measure and Report: H-1; M-14; L-2; I-0
- Reliability: H-0; M-8; L-7; I-0
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF Scientific Methods Panel's ratings for Reliability: High/Moderate/Low (H-3; M-3; L-1; I-1)
- Validity: H-0; M-8; L-7; I-0
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF Scientific Methods Panel's ratings for Reliability: High/Moderate/Low (H-3; M-3; L-1; I-1)
- Feasibility: H-6; M-11; L-0; I-0
- Use: Pass-16; No Pass-1
- Usability: H-0; M-11; L-6; I-0

Standing Committee Recommendation for Endorsement: Vote not taken

The Standing Committee did not vote on the recommendation for endorsement at the meeting because the Committee did not reach consensus on reliability and validity—must-pass criteria. The Committee will revote on the measure during the post-comment web meeting on October 1, 2020.

The Committee acknowledged that the measure was developed to address the resource use aspect of the Improving Medicare Post-Acute Care Transformation Act of 2014 (IMPACT Act). The Committee reviewed a range of data demonstrating high impact through differences in PAC payments across home health agencies (HHAs). It agreed that the range of performance demonstrated an opportunity for improvement in reducing the variability in spending and passed the measure on the importance of the measure and report criterion.

The Committee noted that this measure has been evaluated by the SMP, which gave passing ratings for both reliability and validity. For reliability, the Committee reviewed the signal-to-noise (S/N) analysis and split-sample reliability testing conducted by the developer. Some Committee members raised concerns with some of the lower quartile reliability scores for the taxpayer identification number-national provider identifier (TIN-NPI) reporting level (0.63 and 0.57 for the S/N and split-sample, respectively), stating that they are low. Several Committee members noted that it may be difficult to differentiate HHAs with smaller number of qualifying episodes. Due to these concerns the Committee did not reach consensus on reliability.

For validity, the Committee questioned whether HHAs would be able to control costs that resulted after their care and questioned the developer's decision to utilize a 60-day episode period. The developer clarified that as the measure emphasized upstream interventions and coordination of care, the costs associated with the amount of care needed during hospitalization or emergency department use can be

influenced by HHAs. The developer further clarified that though HHA care tended to be long term, and that the first 60 days of HHA care is a strong indicator of downstream outcomes. The Committee also had concerns that the calculation of expected cost was not aligned with HHA payment programs The developer noted on the call that there is more alignment in the expected episode cost and the payment program for LTCH, and less for the HHA and the other PAC measures. Lastly, the Committee questioned how well the model predicts downstream costs (r-squared of 0.092) and raised concerns regarding the lack of including social factors. The developer indicated that though each of the social factors tested was statistically significant in the model, they did not improve the model fit. Additionally, the developer commented that the dual eligibility in the social risk factor testing carries a negative coefficient, which would lower expected cost. This would penalize providers for taking care of dual-eligible beneficiaries in certain episodes. The Committee stressed that inclusion of risk factors should be about minimizing bias and may not always improve model fit. Due to these concerns, the Committee did not reach consensus on validity.

The Committee agreed that this measure would be feasible as all were routinely collected and posed no additional data collection burden on providers The Committee passed this measure on use and usability, acknowledging that this is a new measure and that it is publicly reported as part of the Centers of Medicare & Medicaid Services' Home Health Quality Reporting Program.

3574 Medicare Spending Per Beneficiary (MSBP) Clinician [Centers for Medicare and Medicaid/Acumen, LLC]

Measure Steward/Developer Representatives at the Meeting

Nirmal Choradia, Acumen LLC
Alan Levitt, Centers for Medicare and Medicaid Services (CMS)
Ronique Evans, Centers for Medicare and Medicaid Services (CMS)

Standing Committee Votes

- Importance to Measure and Report: H-3; M-13; L-1; I-0
- Reliability: H-0; M-9; L-6; I-0
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF Scientific Methods Panel's ratings for Reliability: High/Moderate/Low (H-1; M-4; L-3; I-0)
- Validity: H-0; M-5; L-10; I-0
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.
    - The NQF Scientific Methods Panel's ratings for Reliability: High/Moderate/Low (H-0; M-5; L-3; I-0)
- Feasibility: Vote not taken
- Use: Vote not taken
- Usability: Vote not taken

Standing Committee Recommendation for Endorsement: Vote not taken

The Standing Committee did not recommend the measure for initial endorsement because the Committee did not pass the measure on validity—a must-pass criterion.

The Committee acknowledged the 2017 MedPAC report cited by the developer indicating that inpatient hospital spending accounted for 22% of total Medicare spending in 2015 and represented the second largest Medicare spending category in 2015. The Committee reviewed data provided by the developer demonstrating that MPSB episodes have a range of cost performance at the taxpayer identification number (TIN) level and the TIN-national provider identifier (TIN-NPI) level. The Committee agreed that there is an opportunity for improvement and ultimately passed the measure on the importance to measure and report criterion.

The Committee noted that this measure has been evaluated by the SMP, which gave passing ratings for both reliability and validity. For reliability, the Committee reviewed the signal-to-noise analysis and split sample reliability testing conducted by the developer. Some Committee members raised concerns with some of the lower quartile reliability scores for the TIN-NPI reporting level (0.60), stating that they are low. The developer noted that this may be due to low participation in the CMS Merit-based Incentive Payment System for TIN-NPI, which declined from 2017 to 2018. However, given this concern by some of the Committee members, the Committee did not reach consensus on reliability.

For validity, the Committee reviewed the results for both face and empirical validity testing conducted by the developer. The Committee did not have any concerns regarding the face validity but did raise several concerns regarding the empirical validity of the measure. Some members questioned the attribution to multiple clinicians and whether a care episode could be attributed to multiple clinician groups and multiple clinicians. The Committee also questioned the validity of the time window of three and 30 days pre- and post-discharge for each episode, respectively, and that this might need to be more specific for certain medical conditions. It also questioned the strength of the correlations, noting that the correlation between predicted value and the six different clinical themes (e.g., PAC settings) and the correlation with the risk-adjusted value and the six different clinical themes were low. Lastly, the Committee questioned how well the model predicts downstream costs after a hospitalization and raised concerns regarding the lack of including social factors. The developer noted that testing demonstrated significance of the social factors, but inconsistent direction of the social risk factors and limited impact of social risk factor effects under the current risk adjustment model. The Committee noted that risk adjustment should be focused on reducing bias and may not improving the model fit. The Committee ultimately did not pass the measure on validity.

3575 Total Per Capita Cost (TPCC) [Centers for Medicare and Medicaid/Acumen, LLC]

Measure Steward/Developer Representatives at the Meeting

Nirmal Choradia, Acumen LLC
Alan Levitt, Centers for Medicare and Medicaid Services (CMS)
Ronique Evans, Centers for Medicare and Medicaid Services (CMS)

Standing Committee Votes

- Importance to Measure and Report: H-6; M-11; L-0; I-0

- Reliability: H-0; M-14; L-1; I-0

    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.

  - The NQF Scientific Methods Panel's ratings for Reliability: High/Moderate/Low (H-1; M-6; L-0; I-0)
  - The Committee did not accept the NQF Scientific Methods Panel's Moderate rating: Yes-X; No-X

- Validity: H-0; M-9; L-6; I-0
  - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel.
  - The NQF Scientific Methods Panel's ratings for Reliability: High/Moderate/Low (H-1; M-4; L-2; I-0)
- Feasibility: H-6; M-11; L-0; I-0
- Use: Pass-16; No Pass-1
- Usability: H-0; M-16; L-1; I-0

Standing Committee Recommendation for Endorsement: Vote not taken

The Standing Committee did not vote on the recommendation for endorsement at the meeting because the Committee did not reach consensus on validity—a must-pass criterion. The Committee will revote on the measure on the post-comment web meeting on October 1, 2020.

The Committee acknowledged research cited by the developer that indicated how primary care management in certain settings, such as patient-centered medical homes, can reduce the total cost of care by reducing utilization of high-cost service. It reviewed data provided by the developer demonstrating that total per capita cost (TPCC) has a range of cost performance at the TIN level and the TIN-NPI level. The Committee agreed that there is an opportunity for improvement and ultimately passed the measure on the Importance to Measure and Report criterion.

The Committee noted that this measure has been evaluated by the SMP, which gave passing ratings for both reliability and validity. For reliability, it reviewed the signal-to-noise analysis and split sample reliability testing conducted by the developer. The Committee did not raise any major concerns and agreed with the SMP that the measure was reliable and passed the measure on this criterion.

For validity, the Committee reviewed the results for both face and empirical validity testing conducted by the developer. The Committee did not have any concerns regarding the face validity but did raise several concerns regarding the empirical validity of the measure. It questioned the strength of the correlations, noting that the correlation with risk- and specialty-adjusted cost were low to moderate. The Committee also raised concerns regarding the lack of social factors in the risk adjustment model. The developer reported that inclusion of social factors in the model did not significantly change TIN or TIN-NPI performance scores on average. The Committee noted that risk adjustment should be focused on reducing bias and may not improving the model fit. The Committee ultimately did not reach consensus on validity.

Moving to feasibility, the Committee agreed that this measure would be feasible as all were routinely collected and posed no additional data collection burden on providers. The Committee also passed this measure on use and usability. They acknowledged that the measure is publicly reported as part of the Centers of Medicare & Medicaid Services' Quality Payment Program Merit-based Incentive Payment

System (MIPS) and this measure will be implemented as part of MIPS beginning in the 2020 MIPS performance year and 2022 MIPS payment year.

Public Comment

No public or NQF member comments were provided during the measure evaluation meeting.

Next Steps

NQF will post the draft technical report on August 14, 2020 for public comment for 30 calendar days. The continuous public comment with member support will close on September 14, 2020. NQF will reconvene the Standing Committee for the post-comment web meeting on October 1, 2020.