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 three web meetings on July 9, 13, and 27, 2021 to evaluate five measures undergoing maintenance review.

Welcome, Introductions, and Review of Meeting Objectives

Dr. Matthew Pickering, NQF senior 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. One Cost and Efficiency Standing Committee member, Bijan Borah, MSc, PhD, was recused from voting on the scientific acceptability criteria (i.e., reliability and validity) for all five measures due to his involvement with the NQF Scientific Methods Panel (SMP), which had previously voted on these criteria. Lastly, three Standing Committee members, Dinesh Kalra, Suman Majumdar and John Brooks, were termed inactive for this cycle due to competing engagements.

During all three meetings, the quorum (13 out of 19 active Standing Committee members) required for voting was not achieved. Therefore, the Standing Committee discussed all relevant criteria and voted after the meeting using an online voting tool. The meeting recording and transcript were shared with the Standing Committee after each of the three meetings.

Topic Area Introduction and Overview of Evaluation Process

Ms. Janaki Panchal, NQF manager, provided an overview of the topic area and the current NQF portfolio of endorsed measures. There are currently 13 endorsed measures in the Cost and Efficiency portfolio. Additionally, Ms. Panchal reviewed the Consensus Development Process (CDP) and the measure evaluation criteria.

A measure is recommended for endorsement by the Standing Committee when the vote margin on all must-pass criteria (i.e., Importance, Scientific Acceptability, and Use [maintenance measures only]), and overall, is greater than 60 percent of eligible Standing Committee members' vote in favor of endorsement. A measure is not recommended for endorsement when the vote margin on any must-pass criterion or overall is less than 40 percent of Standing Committee members vote in favor of endorsement. The Standing Committee has not reached consensus if the vote margin on any must-pass criterion or overall is between 40 and 60 percent, inclusive, in favor of endorsement. These measures are termed as Consensus Not Reached (CNR) measures. When all the measures for which, including the CNR measures, are released for NQF member and public comment, the measures where the Standing Committee has not reached consensus will be flagged as CNR measures. The Standing Committee will consider the comments and re-vote on those measures during a webinar convened after the commenting period closes.

Measure Evaluation

During the measure evaluation meetings, the Cost and Efficiency Standing Committee evaluated five maintenance measures for continued endorsement. NQF accepts comments for four weeks prior to the measure evaluation meeting. For this evaluation cycle, the commenting period opened on May 6, 2021.

No comments were submitted by the pre-meeting deadline (June 17, 2021). A summary of the Standing Committee's deliberations below will also be compiled and provided in the draft technical report. NQF will post the draft technical report on August 27, 2021, for public comment on the NQF website. The draft technical report will be posted for 30 calendar days.

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

NQF #1598 Total Resource Use Population-Based per Member per Month (PMPM) Index (HealthPartners)

HealthPartners (Measure Steward/Developer) Representatives at the Meeting

- Chad Heim, vice president Health Informatics
- Gary Kitching, senior director, Health Informatics
- Erika Vetta, senior manager, Health Informatics

Description: The Resource Use Index (RUI) is a risk adjusted measure of the frequency and intensity of services utilized to manage a provider group's patients. Resource use includes all resources associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology and ancillary and behavioral health services. An RUI, when viewed together with the Total Cost of Care measure (NQF-endorsed #1604), provides a more complete picture of population-based drivers of health care costs. **Measure Type**: Cost/Resource Use; **Level of Analysis**: Population: Community, County or City, Clinician: Group/Practice; **Setting of Care**: Emergency Department and Services, Home Care, Inpatient/Hospital, Other, Outpatient Services, Post-Acute Care; **Data Source**: Claims

Standing Committee Votes

- **Importance to Measure and Report**: H-7; M-8; L-0; I-0 (denominator = 15)
- **Reliability**: H-7; M-7; L-0; I-0 (denominator = 14; due to SMP member, Bijan Borah, recusal)
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel (SMP).
    - The NQF SMP's rating for Reliability: H-4; M-3; L-0; I-2
    - 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**: H-5; M-7; L-2; I-0 (denominator = 14; due to SMP member, Bijan Borah, recusal)
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel (SMP).
    - The NQF SMP's ratings for Validity: H-4; M-2; L-1; I-2
    - 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.
- **Feasibility**: H-6; M-9; L-0; I-0 (denominator = 15)
- **Use**: Pass-15; No Pass-0 (denominator = 15)
- **Usability**: H-2; M-13; L-0; I-0 (denominator = 15)

Standing Committee Recommendation for Endorsement: Yes-15; No-0

The Standing Committee recommended the measure for continued endorsement.

This measure was discussed during the initial measure evaluation web meeting on July 9, 2021. Originally endorsed in 2017, NQF #1598 is a risk-adjusted measure that focuses on the frequency and intensity of services utilized to manage a provider group's patients. In discussing the importance to measure and report criterion, some Standing Committee members raised concerns with the scope of improvement for this measure. The Standing Committee emphasized that lower cost does not necessarily correlate to improved quality. The Standing Committee noted that if cost and quality are not highly correlated, then the potential for unintended consequences could lower quality of care and asked the developer how they would address this challenge to prevent inadvertently reducing the quality of care. In response, the developer explained that this measure is used to optimize health and patient experience while improving affordability. To that regard, the goal when implementing both NQF #1598 and NQF #1604 is to use quality and resource together so that both criteria improve. The developer emphasized that the impact on quality of care is the greatest when NQF #1598, a resource use measure, and NQF #1604, a cost measure, are implemented together.

One Standing Committee member desired further understanding of the type of distribution represented by the improvement data provided by the developer, specifically the meaningful difference in performance data included in the developer's submission that noted "26 providers were better than average, three were 10 percent better than average, 12 were 10 percent higher than average, and 50 providers were within 10 percent of the average." The developer clarified that the data demonstrated a normal distribution (i.e., slightly skewed towards the higher side) with variation in performance among the total providers included in the measure over time. A Standing Committee member raised the following concern: Based on the measure submission document, the measure aims to make resources comparable across settings, which indicates that the location where the service was provided is irrelevant; however, the reimbursement differs based on where the service was provided, and the total care relative resource value (TCRRV) table does not allow one to ascertain where the service was provided. The developer explained that the difference in the setting will not appear in the current total resource use measure (NQF #1598); however, it will appear in the total cost measure (NQF #1604). The developer emphasized that the two measures should be used together for a comprehensive view on how cost, resource use, and price interact with one another. The Standing Committee observed that this measure gives a top-line indication of resource use; as a result, the health systems and provider groups would need to get involved to determine which opportunity for improvement to focus on and to put interventions in place for improvement. Ultimately, the Standing Committee agreed that this measure addresses a high-impact/high-resource use area of healthcare.

The Standing Committee noted that the SMP evaluated and rated this measure as "high" for both reliability and validity. In evaluating reliability, the Standing Committee noted the SMP's concerns regarding the intended testing of the measure on the HealthPartners data set; subsequently, the testing would not be generalizable to other practices. The Standing Committee then asked the developers to clarify whether they have tested the measure on any other data set outside of the HealthPartners data set. The developers explained that HealthPartners has a network of providers, and they only have access to those data. Therefore, they do not have access to the data for providers outside of their network. Other organizations have implemented and are using the measure nationally; however, they also do not have access to those data. The developers added that the testing is broad enough for the measure to be potentially used to infer its effectiveness in other markets, as seen with the other organizations using the measure, such as the Network for Regional Healthcare Improvement (NRHI) in California, Minnesota (MN) Community Measurement, and other organizations. The developers elaborated that MN Community Measurement releases a report each year that combines data from four different plans across MN, and in comparing the yearly trends, MN Community Measurement has observed improvement each year. In reviewing the testing methods, both the SMP and the Standing Committee

noted that both the bootstrap and 90 percent random sampling methods theoretically work for a large sample; however, it is unclear how the results would change when applied to smaller provider groups (i.e., providers with less than 600 members). The developers agreed that lower provider sizes indicate higher variation; nonetheless, this measure can be used with lower provider sizes and can still produce reliable results due to the removal of outliers, which can drastically affect the average value. The Standing Committee did not raise any further questions or concerns regarding reliability and passed the measure on this criterion.

Regarding validity, one Standing Committee member questioned whether the TCRRV incorporates pricing. The developer clarified that the TCRRVs are constructed in a way that removes price from the methodology so that the average paid amount is standardized across all hospital providers. Another Standing Committee member noted the appropriateness of presenting correlation coefficients with adjusted clinical group (ACG) scores; however, they raised concerns with how to interpret the correlation coefficient with non-risk-adjusted per member per month capitation payments (PMPMs) because the model is, in fact, risk-adjusted. The developer explained that when the measure is not risk-adjusted, the correlations are high between the PMPMs and the TCRRVs. The developer is attempting to draw a parallel between some known healthcare utilization markers to demonstrate that TCRRVs are a good reflection of resource use. The developers further explained that they are comparing the non-risk adjusted PMPM to the ACG correlation coefficients, because ACGs are a proven measure of resources; and by showing that TCRRVs align well with ACGs, the developers demonstrate that TCRRVs are also a good metric for resource use. Some Standing Committee members also raised concern about excluding patients over the age of 64 and questioned the rationale for using the truncation level of 125,000 TCRRVs. The developer noted that members over the age of 64 are excluded due to potential incomplete claims data of a Medicare-eligible beneficiary. Sometimes, the certain members possess dual coverage, meaning they are in both Medicare and commercial plans; therefore, due to the difficulty in parsing out the costs that are paid for by Medicare versus the commercial plan, members over the age of 64 were excluded. In terms of truncation, the developer noted that the truncation level was selected to remain aligned with healthcare costs and ensure stability of the measure. TCRRVs per member above $125,000 are also excluded (i.e., truncated). One of the Standing Committee members also expressed a need for a different approach to looking at social risk factors because the testing data do not match the reality that both race and income have an impact on the measured outcome. It was noted that testing the addition of just one or two socio-economic status (SES) factors obscures the impact on the measured outcome due to partial effects for other variables in the model (e.g., comorbidities), and stratifying results by SES factors would likely reveal more disparities. As NQF's technical guidance on risk adjustment becomes finalized, the developer will take the previously stated information regarding SES factors into consideration when evolving their criteria for risk adjustment and including SES factors in the risk adjustment model. Without any further concerns, the Standing Committee passed the measure on the validity criterion.

Moving to feasibility, some Standing Committee members expressed concern that the implementation of the measure requires the use of ACGs, which must be licensed separately and may be cost-prohibitive for some entities. Although the ACG is proprietary, the developer has organized a public-facing website with several resources and technical documentation, including toolkits for external organizations to download the necessary tools to run the measure, free of charge. The developer has created instructions and toolkits for both statistical analysis system (SAS) and non-SAS users. The Standing Committee did not raise any further concerns and passed the measure on feasibility. While discussing the usability and use criteria, some Standing Committee members expressed general concerns about the specifications not being fully transparent and noted that it was unclear as to how the measure was being reported and used. Some Standing Committee members also questioned how the feedback on quality

improvement was being provided to the provider groups and asked the developers to clarify what type of feedback was being provided. The developer clarified that they provide quarterly comprehensive reports and monthly patient applications to best support providers in identifying opportunities for improving affordability for their patients, while simultaneously supporting patient outreach, pre-visit planning, and care coordination efforts. The developer also engages with their provider groups and network on an ongoing basis and organized a public-facing website with several resources and technical documentation, including toolkits for external organizations to download the necessary tools to run the measure, free of charge. One Standing Committee member whether there are any data on whether this measure causes any harm to patients. The developer confirmed that they have not received any feedback on any harm caused by this measure. The Standing Committee observed that this measure has been used as a quality improvement tool within the HealthPartners network, and the developer has shown that it has been used by external organizations; the developer also noted they are collecting feedback on this measure. The Standing Committee did not raise any additional questions or concerns, passed the measure on usability and use, and recommended this measure for continued endorsement.

NQF #1604 Total Cost of Care Population-Based PMPM Index (HealthPartners)

Measure Steward/Developer Representatives at the Meeting

- Chad Heim
- Gary Kitching
- Erika Vetta

Description: The Total Cost of Care reflects a mix of complicated factors such as patient illness burden, service utilization and negotiated prices. The Total Cost Index (TCI) is a measure of a primary care provider's risk adjusted cost effectiveness at managing the population they care for. The TCI includes all costs associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, and ancillary and behavioral health services. A TCI, when viewed together with the Total Resource Use measure (NQF-endorsed #1598) provides a more complete picture of population-based drivers of health care costs. **Measure Type**: Cost/Resource Use; **Level of Analysis**: Population: Community, County or City, Clinician: Group/Practice; **Setting of Care**: Emergency Department and Services, Home Care, Inpatient/Hospital, Other, Outpatient Services, Post-Acute Care; **Data Source**: Claims

Standing Committee Votes

- **Importance to Measure and Report**: H-7; M-8; L-0; I-0 (denominator = 15)
- **Reliability**: H-8; M-6; L-0; I-0 (denominator = 14; due to SMP member, Bijan Borah, recusal)
    - This measure is deemed as complex and was evaluated by the SMP.
    - The NQF SMP's rating for Reliability: H-4; M-3; L-0; I-2
    - 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**: H-3; M-8; L-3; I-0 (denominator = 14; due to SMP member, Bijan Borah, recusal)
    - This measure is deemed as complex and was evaluated by the SMP.
    - The NQF SMP's rating for Validity: H-4; M-2; L-1; I-2

    - 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**: H-7; M-7; L-1; I-0
- **Use**: Pass-15; No Pass-0
- **Usability**: H-5; M-8; L-2; I-0

Standing Committee Recommendation for Endorsement: Yes-14; No-1

The Standing Committee recommended the measure for continued endorsement.

This measure was discussed during the initial measure evaluation web meeting on July 9, 2021. This measure was originally endorsed in 2017 and captures the primary care provider's risk-adjusted cost of managing the population they care for. The Standing Committee observed that this measure is very similar to NQF #1598, and the concerns raised for that measure also apply to this measure (NQF #1604). The developer also noted the similarity in the two measures and added that the difference between the two measures is in the costing approach. NQF #1598 measures resource use for every service, while NQF #1604 utilizes an allowed amount, which reflects both the plan and member liabilities. In comparing the two measures, one Standing Committee member questioned how the incorporation of reimbursements in this measure affects providers who practice in higher cost of living and higher geographic wage areas versus lower cost of living and lower geographic wage areas. The developer clarified that the cost differential is driven by the price differential, and this measure is able to show the cost variation that is occurring across different states. One Standing Committee member further questioned whether the developer was conducting stratification to confirm that the differences in cost across providers are due to practice style and not the geographic differences in the labor wage rates. The developer confirmed that this is how the measure is applied so that the comparison is made among the providers in the same wage market. The Standing Committee noted that the other discussed concerns related to the importance to measure and report criterion for NQF #1598 also apply to this measure. They agreed that this measure addresses a high-impact/high-resource use area of healthcare and passed the measure on this criterion.

The Standing Committee noted that the SMP evaluated and rated this measure as high for both reliability and validity. The Standing Committee once again emphasized that the concerns regarding the reliability and validity of this measure are similar to those raised for NQF #1598. The Standing Committee requested additional documentation from the developers regarding the signal-to-noise analysis, which the developer offered to share after the meeting for the Standing Committee's reference. One Standing Committee member asked how a patient is attributed to different clinical practices, specifically what the attribution rules are when multimorbid patients are being handled by multiple providers. The developer explained that they used 12 months of claims data to identify the primary care provider that the member visited most frequently. The Standing Committee did not have any further questions or concerns and passed the measure on reliability and validity.

Regarding feasibility, the Standing Committee observed that there are barriers to access because the ACG system is proprietary; however, this has been addressed by the training, software, and data collection provided by the developer. Therefore, the Standing Committee passed the measure on feasibility. In discussing usability and use, the Standing Committee questioned whether the performance results are available to outside organizations or practices whose performance is being measured. The developers once again emphasized that most of the external organizations publish performance reports publicly, while others do not. The Standing Committee acknowledged that the concerns for usability and

use are similar to those for NQF #1598. They passed the measure on usability and use and recommended the measure for continued endorsement.

NQF # #2431 Hospital-Level, Risk-Standardized Payment Associated With a 30-Day Episode-of-Care for Acute Myocardial Infarction (AMI) (Centers for Medicare & Medicaid Services [CMS]/Yale CORE)

Description: This measure estimates the hospital-level, risk-standardized payment for an acute myocardial infarction (AMI) episode-of-care, starting with inpatient admission to a short-term, acute-care facility and extending 30 days post-admission for Medicare fee-for-service (FFS) patients who are 65 years of age or older with a principal discharge diagnosis of AMI. **Measure Type**: Cost/Resource Use; **Level of Analysis**: Facility; **Setting of Care**: Inpatient/Hospital; **Data Source**: Claims, Enrollment Data

Measure Steward (CMS)/Developer (Yale CORE) Representatives at the Meeting

- Sapha Hassan
- Doris Peter
- Jacqueline Grady
- Smitha Vellanky
- Lisa Suter
- Huihui Yu

Standing Committee Votes

- **Importance to Measure and Report**: H-5; M-9; L-1; I-0 (denominator = 15)
- **Reliability**: H-5; M-6; L-3; I-0 (denominator = 14; due to SMP member, Bijan Borah, recusal)
    - This measure is deemed as complex and was evaluated by the SMP.
    - The NQF SMP's rating for Reliability: H-3, M-5, 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**: H-3; M-7; L-3; I-1 (denominator = 14; due to SMP member, Bijan Borah, recusal)
    - This measure is deemed as complex and was evaluated by the NQF Scientific Methods Panel (SMP)
    - The NQF SMP's rating for Validity: H-1, 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**: H-8; M-7; L-0; I-0
- **Use**: Pass-15; No Pass-0
- **Usability**: H-1; M-10; L-3; I-1

Standing Committee Recommendation for Endorsement: Yes-11; No-4

The Standing Committee recommended the measure for continued endorsement.

This measure was previously endorsed in 2015. The Standing Committee acknowledged that AMI is one of the leading causes of hospitalization for Americans over 65 years of age and costs the United States (U.S.) roughly $84.9 billion in direct and indirect costs. The Standing Committee considered the performance data that the developer reported, which included a mean risk-standardized payment (RSP)

of $25,561 with a range of $17,488 – $32,81 for a reporting period of July 1, 2016-June 30, 2019 and a median hospital RSP in the combined three-year data set of $25,422 (interquartile range of $24,859 – $26,165). The Standing Committee discussed whether this measure is linked to quality. The developer mentioned that this measure is intended to be used in conjunction with other outcome measures, particularly beyond the mortality measure, which would be indicators of quality. The Standing Committee considered whether it would be a hospital's responsibility to find the right quality score or whether the measure results would be presented in a manner that will help hospitals manage and reduce the variation. The developer responded confirmed that this measure is intended to be reported together with other quality measures, such as mortality. The measure results would be reported to the hospitals, including how hospitals compared with other hospitals, in this case for the mortality measure. The developer also mentioned that this measure is not in a pay-for-performance program. Rather, it is used in a pay-for-reporting program, in which the broader goal is to make hospitals health-conscious; in addition, there is no penalty for being above or below the national average. Hospitals are rewarded if they report on this measure. The Centers for Medicare and Medicaid Services (CMS) explained that this measure and several other condition-specific cost measures are designed to promote transparency as well as target cost reductions. CMS has a Medicare Spending per Beneficiary (MSPB) measure that covers a much broader list of conditions and to which payments are made based on the performance of that measure. The Standing Committee did not raise any further questions and passed the measure on the importance to measure and report criterion.

Moving to scientific acceptability (i.e., reliability and validity), the Standing Committee considered the reliability testing, which was conducted at the performance measure score level. Using the Spearman-Brown prediction formula, the developer found that the agreement between the two independent assessments of the RSP for each hospital was 0.681. The Standing Committee acknowledged that the SMP reviewed and passed this measure on reliability. The Standing Committee asked whether signal-to-noise tests were conducted. In response, the developer explained that hospitals with at least 25 admissions possessed a median signal-to-noise value of 0.404 with an interquartile range of 0.298 to 0.594. One Standing Committee member raised concern that the Standing Committee has considered higher cutoffs for some of the other measures rather than 0.4, such as 0.6 or 0.7. It was noted that the SMP is reviewing various thresholds of reliability to determine which ones are acceptable. In addition, CMS has looked at 0.4 as an acceptable threshold due the tradeoffs of trying to include more facilities or providers within the measure to promote more transparency across the system. The Standing Committee did not raise any additional concerns and passed the measure on reliability.

Moving to validity, the Standing Committee considered that the developer conducted face validity testing in which eight of the 16 developer-convened Technical Expert Panel (TEP) members agreed that the measure can discern good versus poor quality of care. The developer also conducted empirical validity testing, comparing this measure to the MSP measure. The developer found a correlation coefficient of 0.281 ($p < 0.0001$), meaning that hospitals with higher spending across all Medicare FFS beneficiaries correlated with hospitals with higher spending on patients hospitalized with AMI. The Standing Committee considered the risk adjustment model, noting that the R-squared value slightly increased up to 0.078, which suggests that approximately 8 percent of the variation in payment could be explained by patient-level risk factors. The Standing Committee discussed the approach to social risk adjustment, noting that testing the addition of just one or two social risk factors obscures the impact due to partial effects for other variables in the model (e.g., comorbidities). Stratifying results by social risk would also likely reveal more disparities. The developer noted that they tested the impact of dual-eligible status and the Agency for Healthcare Research and Quality (AHRQ) SES index as social risk factors. The developer found that the two social risk factors did have a slightly lower payment after adjustment for other risk factors in the multivariate model; however, the addition of these social risk

factors had limited impact on model performance and produced little change in the measure scores, while the measure scores estimated with hospitals both with and without dual eligibility were highly correlated.

NQF staff stated that NQF is currently developing technical guidance, which was out for public comment at the time of this meeting, that will provide more clarity for developers and NQF Standing Committees in conducting social risk factor adjustment within quality measurement. NQF staff further noted that this guidance will not change NQF's evaluation criteria until the guidance is finalized at the end of next year (2022). The Standing Committee did not raise any further concerns and passed the measure on validity.

The Standing Committee did not raise any concerns with feasibility and passed the measure on this criterion. For use and usability, the Standing Committee acknowledged that this measure is currently publicly reported within CMS' Care Compare and used within the Hospital Inpatient Quality Reporting (IQR) Program. In reviewing the usability criterion, the Standing Committee noted the very small differences in hospital-level RSPs for the AMI payment in 2018-2019 compared with the prior individual years (2016/2017; 2017/2018). The median RSPs for each year were $25,248, $25,539, and $25,542 for 2016/2017, 2017/2018, and 2018/2019, respectively. The Standing Committee asked whether any longer-term harms of the measure's use have been identified. In response, the developer explained that they did not identify any unintended consequences during measure development and testing. The Standing Committee did not raise any concerns, passed the measure on use and usability, and ultimately recommended the measure for continued endorsement. The Standing Committee will review related and competing measures during the post-comment call on October 22, 2021.

NQF #2436 Hospital-Level, Risk-Standardized Payment Associated With a 30-Day Episode-of-Care for Heart Failure (HF) (CMS/Yale CORE)

Description: This measure estimates the hospital-level, risk-standardized payment for a heart failure (HF) episode of care, starting with inpatient admission to a short term acute-care facility and extending 30 days post-admission for Medicare fee-for-service (FFS) patients who are 65 years of age or older with a principal discharge diagnosis of HF. **Measure Type**: Cost/Resource Use; **Level of Analysis**: Facility; **Setting of Care**: Inpatient/Hospital; **Data Source**: Claims, Enrollment Data

Measure Steward/Developer Representatives at the Meeting

- Sapha Hassan
- Doris Peter
- Jacqueline Grady
- Smitha Vellanky
- Lisa Suter
- Huihui Yu

Standing Committee Votes

- **Importance to Measure and Report**: H-4; M-9; L-2; I-0 (denominator = 15)
- **Reliability**: H-6; M-8; L-0; I-0 (denominator = 14; due to SMP member, Bijan Borah, recusal)
    - This measure is deemed as complex and was evaluated by the SMP.
    - The NQF SMP's rating for Reliability: H-5, M-3, 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**: H-2; M-9; L-3; I-0 (denominator = 14; due to SMP member, Bijan Borah, recusal)

    - This measure is deemed as complex and was evaluated by the SMP.
    - The NQF SMP's rating for Validity: H-2, M-4, 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**: H-6; M-9; L-0; I-0
- **Use**: Pass-15; No Pass-0
- **Usability**: H-2; M-8; L-4; I-1

Standing Committee Recommendation for Endorsement: Yes-14; No-1

The Standing Committee recommended the measure for continued endorsement.

This measure was previously endorsed in 2015. The Standing Committee acknowledged that HF is one of the top three leading causes of hospitalization for Americans over 65 years of age and is projected to cost the U.S. up to $70 billion in direct and indirect costs by 2030.

The Standing Committee considered the performance data that the developer reported, which included a mean RSP of $17,722 with a range of $13,171 – $27,996 during the reporting period of July 1, 2016-June 30, 2019 and a median hospital RSP in the combined three-year data set of $17,607 (interquartile range of $16,817 – $18,513). The Standing Committee noted that the same concerns raised during the discussion of NQF #2431 apply to this measure as well. The Standing Committee did not raise any further questions and passed the measure on the importance to measure and report criterion.

Moving to scientific acceptability, the Standing Committee considered the reliability testing, which was conducted at the performance measure score level. Using the Spearman-Brown prediction formula, the developer found that the agreement between the two independent assessments of the RSP for each hospital was 0.781. The Standing Committee acknowledged that the SMP reviewed and passed this measure on reliability. The Standing Committee asked about the signal-to-noise results for this measure. In response, the developer explained that hospitals with at least 25 admissions possessed the following HF payment signal-to-noise reliability values: a mean of 0.66, a median of 0.679, and an interquartile range of 0.528 – 0.801. The Standing Committee acknowledged that the same concerns with NQF #2431 apply to this measure as well. The Standing Committee did not raise any additional concerns and passed the measure on reliability.

Moving to validity, the Standing Committee considered that the developer conducted face validity testing, in which eight of the 16 developer-convened TEP members agreed that the measure can discern good versus poor quality of care. The developer also conducted empirical validity testing, comparing this measure to the MSPB measure. The developer found a correlation coefficient of 0.543, meaning that hospitals with higher spending across all Medicare FFS beneficiaries correlated with hospitals with higher spending on patients hospitalized with HF.

The Standing Committee considered the risk adjustment model, noting that the R-squared value slightly decreased to 0.031, which suggests that approximately 3 percent of the variation in payment could be explained by patient-level risk factors. The Standing Committee discussed the approach to social risk adjustment, noting that the developer tested the impact of dual-eligible status and the AHRQ SES index as social risk factors. The developer found that the two social risk factors did have a slightly lower payment after adjustment for other risk factors in the multivariate model; however, the addition of these social risk factors had limited impact on model performance and produced little change in measure scores and measure scores estimated with hospitals with and without dual eligibility were

highly correlated. The Standing Committee acknowledged that the same concerns with NQF #2431 apply to this measure. The Standing Committee did not raise any additional concerns and passed the measure on validity.

The Standing Committee did not raise any concerns with feasibility and passed the measure on this criterion. For use and usability, the Standing Committee acknowledged that this measure is currently publicly reported within CMS' Care Compare and used within the Hospital IQR Program. In reviewing the usability criterion, the Standing Committee noted that developer reported a median hospital 30-day RSP of $17,607 for the HF payment measure for the three-year period between July 1, 2016, and June 30, 2019. The median RSP decreased by 2.6 percent from July 2017-June 2018 (median RSP: $17,781) to July 2018-June 2019 (median RSP: $17,310). The Standing Committee acknowledged that the same concerns with NQF #2431 for usability apply to this measure as well. The Standing Committee did not raise any further concerns, passed the measure on use and usability, and ultimately recommended the measure for continued endorsement. The Standing Committee will review related and competing measures during the post-comment call on October 22, 2021.

NQF # 2579 Hospital-Level, Risk-Standardized Payment Associated With a 30-Day Episode of Care for Pneumonia (PN) (CMS/Yale CORE)

Description: This measure estimates the hospital-level, risk-standardized payment for an eligible pneumonia (PN) episode of care, starting with inpatient admission to a short-term, acute-care facility and extending 30 days post-admission for Medicare fee-for-service (FFS) patients who are 65 years or older with a principal discharge diagnosis of PN or principal discharge diagnosis of sepsis (not including severe sepsis) that have a secondary discharge diagnosis of PN coded as present on admission (POA) and no secondary diagnosis of severe sepsis coded as POA. **Measure Type**: Cost/Resource Use; **Level of Analysis**: Facility; **Setting of Care**: Inpatient/Hospital; **Data Source**: Claims, Enrollment Data

Measure Steward/Developer Representatives at the Meeting

- Sapha Hassan
- Doris Peter
- Jacqueline Grady
- Smitha Vellanky
- Lisa Suter
- Huihui Yu

Standing Committee Votes

- **Importance to Measure and Report**: H-3; M-10; L-2; I-0 (denominator = 15)
- **Reliability**: H-5; M-9; L-0; I-0 (denominator = 14; due to SMP member, Bijan Borah, recusal)
    - This measure is deemed as complex and was evaluated by the SMP.
    - The NQF SMP's rating for Reliability: H-5, M-3, 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**: H-3; M-9; L-1; I-1 (denominator = 14; due to SMP member, Bijan Borah, recusal)
    - This measure is deemed as complex and was evaluated by the SMP.
    - The NQF SMP's rating for Validity: H-2, M-4, 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**: H-7; M-8; L-0; I-0
- **Use**: Pass-15; No Pass-0
- **Usability**: H-2; M-9; L-4; I-0

Standing Committee Recommendation for Endorsement: Yes-15; No-0

The Standing Committee recommended the measure for continued endorsement.

This measure was previously endorsed in 2014. The Standing Committee acknowledged that pneumonia is one of the top three leading causes of hospitalization for Americans over 65 years of age and costs the U.S. approximately $13.4 billion annually.

The Standing Committee considered the performance data that the developer reported, which included a mean RSP of $18,283 with a range of $10,529 – $29,861 during the reporting period of July 1, 2016-June 30, 2019 and a median hospital RSP in the combined three-year data set of $18,200 (interquartile range of $17,015 – $19,453). The Standing Committee noted that the same concerns raised during the discussion of NQF #2431 apply to this measure as well. The Standing Committee did not raise any further questions and passed the measure on the importance to measure and report criterion.

In reviewing scientific acceptability, the Standing Committee considered the reliability testing, which was conducted at the performance measure score level. Using the Spearman-Brown prediction formula, the agreement between the two independent assessments of the RSP for each hospital was 0.815. The Standing Committee acknowledged that the SMP reviewed and passed this measure on reliability. The Standing Committee asked about the signal-to-noise results for this measure. In response, the developer explained that hospitals with at least 25 admissions had mean signal-to-noise value of 0.820, a median of 0.8554, and an interquartile range of 0.7472 to 0.919. The Standing Committee acknowledged that the same concerns with NQF #2431 and NQF #2436 apply to this measure as well. The Standing Committee did not raise any additional concerns and passed the measure on reliability.

Regarding validity, the Standing Committee considered that the developer conducted face validity testing in which 10 of the 16 developer-convened TEP members agreed that the measure can discern good versus poor quality of care. The developer also conducted empirical validity testing comparing this measure to the MSPB measure. The developer found a correlation between PN RSPs and the MSPB score of 0.588, suggesting that hospitals with higher performance on this measure are more likely to have higher MSPB measure performance scores.

The Standing Committee considered the risk adjustment model, noting that the updated and calculated R-squared results was 0.076. The Standing Committee discussed the approach to social risk adjustment, noting that the developer presents analyses that show a significant association between dual eligibility (but not low AHRQ SES) and higher payments, even after adjusting for other risk factors in a multivariable model. Adding the social risk variables results had little impact on model performance and produced little change in measure scores, while the measure scores estimated for hospitals both with and without dual eligibility were highly correlated (0.999). The developer noted that CMS ultimately decided to not adjust this measure for either dual eligibility or the AHRQ SES Index. The Standing Committee acknowledged that the same concerns with NQF #2431 and #2436 apply to this measure as well. The Standing Committee did not raise any additional concerns and passed the measure on validity.

The Standing Committee did not raise any concerns with feasibility and passed the measure on this criterion. For use and usability, the Standing Committee acknowledged that this measure is currently publicly reported within CMS' Care Compare and used within the Hospital IQR Program. In reviewing the usability criterion, the Standing Committee noted that developer reported a median hospital 30-day RSP of $18,200 for the PN payment measure for the three-year period between July 1, 2016 and June 30, 2019. The median RSP decreased by 1 percent from July 2017-June 2018 (median RSP: $18,226) to July 2018-June 2019 (median RSP: $18,037). The Standing Committee acknowledged that the same concerns with NQF #2431 and #2436 for usability apply to this measure as well. The Standing Committee did not raise any further concerns, passed the measure on use and usability, and ultimately recommended the measure for continued endorsement. The Standing Committee will review related and competing measures during the post-comment call on October 22, 2021.

Public Comment

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

Next Steps

Ms. Panchal provided the next steps, noting that NQF staff will prepare a draft technical report, which will detail the Cost and Efficiency Standing Committee's discussion and recommendations on five maintenance measures undergoing continued endorsement. NQF will post the draft technical report on August 27, 2021, for public comment for 30 calendar days. The continuous public commenting period with member support will close on September 27, 2021. NQF will reconvene the Standing Committee for the post-comment web meeting on October 22, 2021, to review and discuss public comments received during the commenting period.

Appendix A: Pre-Evaluation Comments

No comments received as of June 17, 2021.