NATIONAL QUALITY FORUM
Driving measurable health improvements together

Memo

October 22, 2021

To: Cost and Efficiency Standing Committee

From: NQF staff

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

Introduction

On September 27, 2021, NQF closed the public commenting period on five maintenance measures submitted for endorsement consideration as part of the spring 2021 Cost and Efficiency Consensus Development Process (CDP). NQF received a total of three comments for three of the five measures during the post-comment public commenting period. All three comments were submitted by the American Medical Association (AMA). NQF staff reviewed all three comments and determined that all comments required a response from the developer and Standing Committee (SC) consideration.

The Standing Committee's recommendations will be reviewed by the Consensus Standards Approval Committee (CSAC) on November 30, 2021. The CSAC will review the Standing Committee's recommendation for each measure submitted for endorsement consideration. All Standing Committee members are encouraged to attend the CSAC meeting.

Purpose of the Call

The Cost and Efficiency Standing Committee post-comment web meeting is scheduled for October 22, 2021, from 2:00pm – 5:00pm ET. The purpose of the post-comment call is to:

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

Standing Committee Actions

Review this briefing memo and draft report.

Review and consider the full text of all comments received and the proposed responses to the post-evaluation comments (see comment narrative and additional documents included with the call materials).

Review the NQF members' expressions of support of the submitted measures.

Provide feedback and input on proposed post-evaluation comment responses.

Conference Call Information

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

Meeting link: https://nqf.webex.com/nqf/j.php?MTID=m73950e433b199ec8e9c80bd3d841de21

Meeting Number: 2344 248 2437
Meeting Password: QMEvent
Join by phone: 1-844-621-3956

Background

Of the five measures reviewed in this cycle, three are condition specific (acute myocardial infarction [AMI], heart failure [HF], and pneumonia) and two are non-condition specific (population-based). On July 9, 13, and 27, 2021, NQF convened a multistakeholder Standing Committee composed of 22 individuals to evaluate five measures undergoing maintenance review for the spring 2021 cycle. The Standing Committee recommended all five measures for continued endorsement.

The Standing Committee recommended the following measures:

- **NQF #1598** Total Resource Use Population-based Per Member Per Month per-member-per-month (PMPM) Index
- **NQF #1604** Total Cost of Care Population-based PMPM Index
- **NQF #2431** Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction
- **NQF #2436** Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for heart failure
- **NQF #2579** Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia

Comments Received

NQF welcomes comments from both NQF members and the public for the measures under review. Comments were accepted using the NQF's Quality Positioning System (QPS) online tool located on the Cost and Efficiency project webpage.

Pre-evaluation Comments

NQF accepted comments prior to the Standing Committee's evaluation of the measures. For this evaluation cycle, the pre-evaluation comment period was opened on May 6, 2021. No comments were received by the pre-meeting deadline on June 17, 2021.

Post-evaluation Comments

The spring 2021 Cost and Efficiency draft technical report was posted on the project webpage for a 30-day public and NQF member comment period on August 27, 2021. During this comment period, NQF received three comments from one NQF member organization.

Member Council	# of Member Organizations Who Commented
Health Professional	1

- NQF included all post-evaluation comments that were received in the comment narrative posted to the Standing Committee SharePoint site. This comment narrative contains the commenter's name, comment, associated measure, topic (if applicable), and —for the post-evaluation comments—draft responses (including measure steward/developer responses) for the Standing Committee's consideration. To facilitate discussion, the post-evaluation comments have been categorized into major themes. Please note that the organization of the comments into major themes is not an attempt to limit Standing Committee discussion.

  Please review the post-evaluation comment narrative in advance of the meeting scheduled for October 22, 2021 and consider the individual comments along with the major themes and the proposed responses to each theme. Although all comments are subject to discussion, the intent is not to discuss each individual comment during the post-comment call. Instead, NQF staff will spend most of the time considering the themes discussed below and the set of comments as a whole. Please note measure stewards/developers were asked to respond where appropriate. NQF staff has proposed draft responses to the comment themes for the Standing Committee to consider.

Comments and Their Disposition

Themed Comments

Three major themes were identified in the post-evaluation comments, as follows:

Reliability and minimum reliability thresholds

Social risk and risk adjustment

Cost and quality measure correlation

Theme 1 - RELIABILITY/MINIMUM RELIABILITY THRESHOLDS

The AMA voices concern with the signal-to-noise ratio value ranges specified in measure #2431 (median-0.404; interquartile range [IQR] 0.298-0.594) and measure #2436 (median- 0.679; IQR 0.528-0.801). While the AMA recognizes that the minimum acceptable threshold accepted by CMS is currently 0.4, they state that the minimum threshold should be set at 0.7.

The AMA voices concern with the signal-to-noise ratio value ranges specified in measure #2431 (median-0.404; IQR 0.298-0.594) and measure #2436 (median- 0.679; IQR 0.528-0.801). While the AMA recognizes that the minimum acceptable threshold accepted by CMS is currently 0.4, they state that the minimum threshold should be set at 0.7.

Measure Steward/Developer Response:

NQF #2431

In our testing attachment, we provide split-sample reliability. To calculate split-sample reliability, we randomly sampled half of patients within each hospital from a three-year measurement period, calculated the measure for each hospital, and repeated the calculation using the second half of patients. Thus, each hospital is measured twice, but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of

these two subsets agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement, we calculated the intra-class correlation coefficient (Shrout & Fleiss, 1979). For this measure, the split-sample reliability for hospitals with at least 25 cases was 0.681 which falls within the thresholds currently under consideration by the Scientific Methods Panel (SMP).

NQF #2436

In our testing attachment, we provided split-sample reliability. To calculate split-sample reliability, we randomly sampled half of patients within each hospital from a three-year measurement period, calculated the measure for each hospital, and repeated the calculation using the second half of patients. Thus, each hospital is measured twice, but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of these two subsets agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement, we calculated the intra-class correlation coefficient (Shrout & Fleiss, 1979). For this measure, the split-sample reliability was 0.781, which falls within the thresholds currently under consideration by the Scientific Methods Panel (SMP).

Proposed Committee Response:

Thank you for your comments. During the Standing Committee initial review of the measures under consideration, concern was raised regarding the signal-to-noise reliability statistics for entities with low case volume. The Standing Committee acknowledged challenges with achieving reliability thresholds for measure score reliability while balancing the trade-off of including more facilities or providers within the measure to promote transparency across the health care system. The Standing Committee also considered the Scientific Methods Panel (SMP)'s decision to pass the measure on reliability and their input on the reliability testing results when voting to recommend these measures for endorsement.

Action Item:

The Standing Committee should review the comment and developer's response and be prepared to discuss them in relation to the measure and its current recommendation status. The Standing Committee should determine whether they agree with the proposed response.

Theme 2 - SOCIAL RISK AND RISK ADJUSTMENT

The AMA raised concern with testing for social risk factors after adjusting for clinical risk factors rather than assessing the combined clinical and social risk factor impact simultaneously. The AMA also questioned the adequacy of the risk model due to the R-squared results specified in measures #2579 (0.076), #2431 (0.078), and #2436 (0.031). Taking the R-squared results into consideration, the commenter expressed concern that measures #2579, #2431, and #2436 do not meet the scientific acceptability criteria.

Measure Steward/Developer Response:

NQF #2431

As noted earlier, the SMP reviewed this measure, including an assessment of the risk model, and rated it high for validity.

Quasi-R2: For a traditional linear model (i.e. ordinary least squares regression) R2 is interpreted as the amount of variation in the observed outcome that is explained by the predictor variables (patient-level risk factors). Generalized linear models (GLMs), however, do not output an R2 that is akin to the R2 of a traditional linear model. In order to provide the NQF Committee with a

statistic that is conceptually similar, we produced a "quasi- R2" by regressing the total payment outcome on the predicted outcome (Jones et al, 2010). Specifically, we regressed the total payment on the payment predicted by the patient-level risk factors. This regression produces a quasi-R2 that indicates the percent of the variation in payment can be explained by patient-level risk factors. The quasi-R2 results are consistent with R2s from other patient-level risk adjustment models for health care payment (Pope et al., 2011). Additional model performance results (predictive ratios, calibration) support the validity of the risk model for this measure.

Social Risk Factors: It is a standard and acceptable practice to test the incremental effects of social risk factors within a clinical risk model, as increased risk from a single social risk factor may be in part or completely explained by a clinical risk factor already in the model.

The payment measures are meant to be reported along with readmission and mortality measures for the same conditions, and those measures, which were recently recommended for re-endorsement, do not include adjustment for social risk factors. Note that the payment measures are not used in a pay-for-performance program.

We do not dispute that there can be differences in unadjusted, observed outcomes based on social risk – our own results presented in the testing attachment show, for example, that for the low AHRQ SES variable, mean observed payments are slightly higher for patients with the social risk factor compared with patients without the social risk factor; for the dual eligibility variable however, observed payments are lower for patients with the social risk factor (note that due to past feedback from NQF, we did not test any race-related variables). We also note that our results presented in the testing attachment show that payment ratios estimated with models that adjust for either social risk factor are significantly lower than one. The question we are trying to address with our analyses is the impact of adjusting for social risk factor on this particular measure score (risk-standardized payment). Our results show that differences in mean payments are very small, and the correlations between risk-standardized payments for models with and without the social risk factors are near 1.

References:

Jones AM. Models for Health Care. Health, Econometrics and Data Group (HEDG) Working Papers. 2010.

Pope, G. C., Kautter, J., Ingber, M. J., Freeman, S., Sekar, R., & Newhart, C. RTI International, (2011). Evaluation of the CMS-HCC risk adjustment model (Final Report). pp.6.

NQF #2436

As noted earlier, the SMP reviewed this measure, including an assessment of the risk model, and rated it high for validity.

Quasi-R2: For a traditional linear model (i.e. ordinary least squares regression), R2 is interpreted as the amount of variation in the observed outcome that is explained by the predictor variables (patient-level risk factors). Generalized linear models (GLMs), however, do not output an R2 that is akin to the R2 of a traditional linear model. In order to provide the NQF Committee with a statistic that is conceptually similar, we produced a "quasi- R2" by regressing the total payment outcome on the predicted outcome (Jones et al, 2010). Specifically, we regressed the total payment on the payment predicted by the patient-level risk factors. This regression produces a quasi-R2 that indicates the percent of the variation in payment can be explained by patient-level risk factors. The quasi-R2 results are consistent with R2s from other patient-level risk adjustment models for health care payment (Pope et al., 2011). Additional model performance results (predictive ratios, calibration) support the validity of the risk model for this measure.

Social Risk Factors: It is a standard and acceptable practice to test the incremental effects of social risk factors within a clinical risk model, as increased risk from a single social risk factor may be in part or completely explained by a clinical risk factor already in the model.

The payment measures are meant to be reported along with readmission and mortality measures for the same conditions, and those measures, which were recently recommended for re-endorsement, do not include adjustment for social risk factors. Note that the payment measures are not used in a pay-for-performance program.

We do not dispute that there are differences in unadjusted, observed outcomes based on social risk – our own results presented in the testing attachment show, for example, that for the dual eligibility variable, mean observed payments are higher for patients with the social risk factor compared with patients without the social risk factor (note that due to past feedback from NQF, we did not test any race-related variables). The question we are trying to address with our analyses is the impact of adjusting for social risk factor on this particular measure score (risk-standardized payment). Our results show that differences in mean payments are very small, and the correlations between adjusted and unadjusted risk-standardized payments are near 1.

In addition, adjusting for social risk factors would likely remove an important hospital-level effect. A 2019 study, described in the testing attachment and authored by the developer, showed that differences in hospital-level payments for heart failure and pneumonia were associated with hospital characteristics independently from patient characteristics (Krumholz et al, 2019). The study design held constant the social determinants of health that were not expected to change between the two admissions and compared the same people at two different hospitals so that behaviors, social context, and demographic characteristics, including race/ethnicity, were the same. The authors compared payments for the same Medicare patient for two admissions for the same condition – one admission to a low-payment hospital and one admission to a high-payment hospital – and found that patients who were admitted to hospitals with the highest payment profiles incurred higher costs than when they were admitted to hospitals with the lowest payment profiles. The findings suggest that variations in payments to hospitals are, at least in part, associated with the hospitals independently of non-time-varying patient characteristics.

References:

Jones AM. Models for Health Care. Health, Econometrics and Data Group (HEDG) Working Papers. 2010.

Krumholz, H. M., Wang, Y., Wang, K., Lin, Z., Bernheim, S. M., Xu, X., Desai, N. R., & Normand, S.T. 2019. Association of Hospital Payment Profiles With Variation in 30-Day Medicare Cost for Inpatients With Heart Failure or Pneumonia. JAMA network open, 2(11), e1915604.

Pope, G. C., Kautter, J., Ingber, M. J., Freeman, S., Sekar, R., & Newhart, C. RTI International, (2011). Evaluation of the CMS-HCC risk adjustment model (Final Report). pp.6.

NQF #2579

As noted earlier, the SMP reviewed this measure, including an assessment of the risk model, and rated it high for validity.

Quasi-R2: For a traditional linear model (i.e. ordinary least squares regression), R2 is interpreted as the amount of variation in the observed outcome that is explained by the predictor variables (patient-level risk factors). Generalized linear models (GLMs), however, do not output an R2 that is akin to the R2 of a traditional linear model. In order to provide the NQF Committee with a statistic that is conceptually similar, we produced a "quasi- R2" by regressing the total payment

outcome on the predicted outcome (Jones et al, 2010). Specifically, we regressed the total payment on the payment predicted by the patient-level risk factors. This regression produces a quasi-R2 that indicates the percent of the variation in payment can be explained by patient-level risk factors. The quasi-R2 results are consistent with R2s from other patient-level risk adjustment models for health care payment (Pope et al., 2011). Additional model performance results (predictive ratios, calibration) support the validity of the risk model for this measure.

Social Risk Factors: It is a standard and acceptable practice to test the incremental effects of social risk factors within a clinical risk model, as increased risk from a single social risk factor may be in part or completely explained by a clinical risk factor already in the model.

The payment measures are meant to be reported along with readmission and mortality measures for the same conditions, and those measures, which were recently recommended for re-endorsement, do not include adjustment for social risk factors. Note that the payment measures are not used in a pay-for-performance program.

We do not dispute that there are differences in unadjusted, observed outcomes based on social risk – our own results presented in the testing attachment show, for example, that for the dual eligibility variable, mean observed payments are higher for patients with the social risk factor compared with patients without the social risk factor (note that due to past feedback from NQF, we did not test any race-related variables). The question we are trying to address with our analysis is the impact of adjusting for social risk factor on this particular measure score (risk-standardized payment). Our results show that differences in mean payments are very small, and the correlations between adjusted and unadjusted risk-standardized payments are near 1.

In addition, adjusting for social risk factors would likely remove an important hospital-level effect. A 2019 study, described in the testing attachment and authored by the developer, showed that differences in hospital-level payments for heart failure and pneumonia were associated with hospital characteristics independently from patient characteristics (Krumholz et al, 2019). The study design held constant the social determinants of health that were not expected to change between the two admissions and compared the same people at two different hospitals so that behaviors, social context, and demographic characteristics, including race/ethnicity, were the same. The authors compared payments for the same Medicare patient for two admissions for the same condition – one admission to a low-payment hospital and one admission to a high-payment hospital – and found that patients who were admitted to hospitals with the highest payment profiles incurred higher costs than when they were admitted to hospitals with the lowest payment profiles. The findings suggest that variations in payments to hospitals are, at least in part, associated with the hospitals independently of non-time-varying patient characteristics.

References:

Jones AM. Models for Health Care. Health, Econometrics and Data Group (HEDG) Working Papers. 2010.

Krumholz, H. M., Wang, Y., Wang, K., Lin, Z., Bernheim, S. M., Xu, X., Desai, N. R., & Normand, S.T. 2019. Association of Hospital Payment Profiles With Variation in 30-Day Medicare Cost for Inpatients With Heart Failure or Pneumonia. JAMA network open, 2(11), e1915604.

Pope, G. C., Kautter, J., Ingber, M. J., Freeman, S., Sekar, R., & Newhart, C. RTI International, (2011). Evaluation of the CMS-HCC risk adjustment model (Final Report). pp.6.

Proposed Committee Response:

Thank you for your comments. The Standing Committee acknowledges the commenter's concern that cost and resource use measure can be influenced by care received in a healthcare setting but also by clinical processes and social risk factors (SRF). While the developers did test for the impact of SRF in the risk models for these measures, some of the measures did not include SRF in the final model. While the Standing Committee notes that it is important to maximize the predictive value of a risk adjustment model, elements of a risk model should be included based on a conceptual and empirical rationale. The Standing Committee considered the SMP input on the validity testing, which passed through SMP, and the approach to the risk adjustment modeling and agreed to recommend these measures for endorsement.

Action Item:

The Standing Committee should review the comment and developer's response and be prepared to discuss them in relation to the measure and its current recommendation status. The Standing Committee should determine whether they agree with the proposed response.

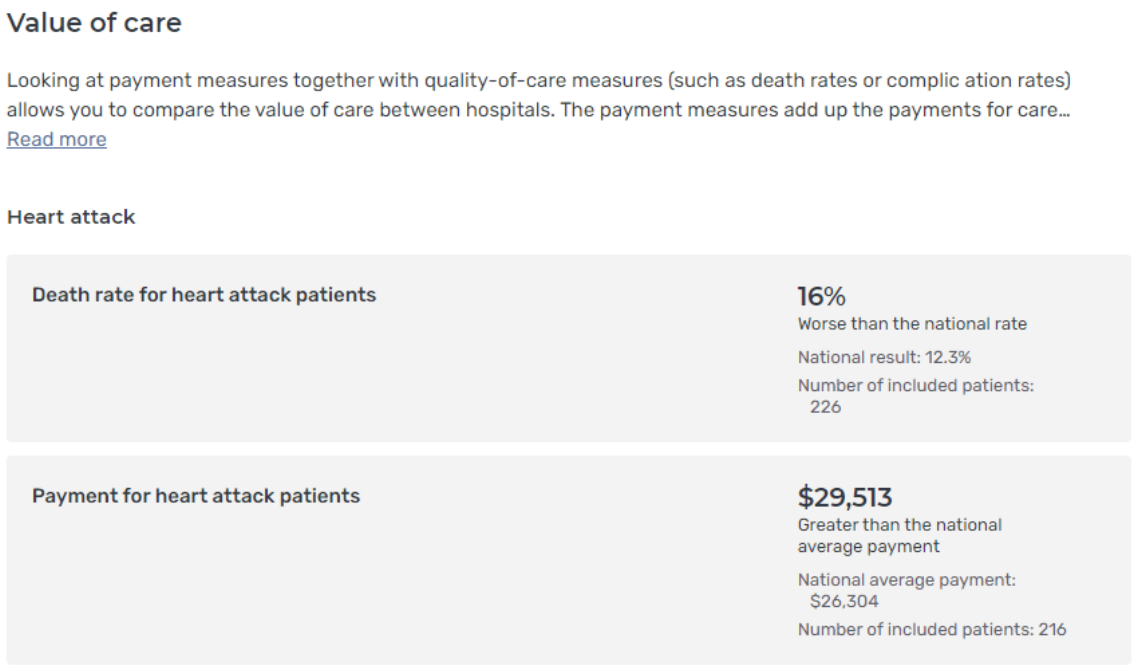
Theme 3 - COST AND QUALITY CORRELATION

The AMA voiced concern with how the developer was unable to demonstrate the correlation between the cost measures (#2431, #2436, and #2579) and any one quality measure within the hospital's quality programs. The commenter noted that the developer mentions that cost measures should not be evaluated alone in the measure specifications.

Measure Steward/Developer Response:

NQF #2431

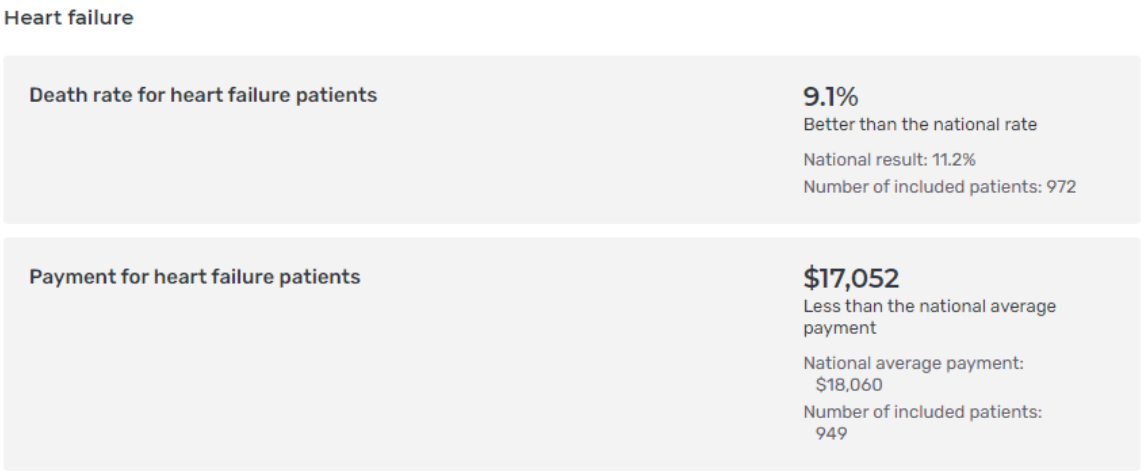
We agree with the AMA that costs need to be assessed within the context of quality of care and have stated so in our submission. Results of the measure alone do not necessarily reflect the quality of care provided by hospitals but simply whether the total episode payments are greater than or less than would be expected for an average hospital with a similar case mix. Accordingly, measure scores are reported together with a quality signal (in this case mortality for the same condition) as an indication of the value of care. (CMS' mortality measures for these conditions were recently re-endorsed by NQF in the Fall 2020 cycle.) An example for one hospital is shown below; this hospital has payments that are greater than the national average and quality that is worse than the national rate, suggesting low-value care.



In addition, each spring, hospitals receive a detailed report of all the patients included in the measure, along with detailed breakdowns of post-acute care costs. Therefore, the payment measures provide an opportunity for hospitals to explore the drivers of costs for their patients and assess the payment measure results in the context of the quality of care they provide to patients.

NQF #2436

We agree with the AMA that costs need to be assessed within the context of quality of care and have stated so in our submission. Results of the measure alone do not necessarily reflect the quality of care provided by hospitals but simply whether the total episode payments are greater than or less than would be expected for an average hospital with a similar case mix. Accordingly, measure scores are reported together with a quality signal (in this case mortality for the same condition) as an indication of the value of care. (CMS' mortality measures for these conditions were recently re-endorsed by NQF in the Fall 2020 cycle.) An example for one hospital is shown below; this hospital has payments that are less than the national average and quality that is better than the national rate, suggesting high-value care.



In addition, each spring, hospitals receive a detailed report of all the patients included in the measure, along with detailed breakdowns of post-acute care costs. Therefore, the payment measures provide an opportunity for hospitals to explore the drivers of costs for their patients and assess the payment measure results in the context of the quality of care they provide to patients.

NQF #2579

We agree with the AMA that costs need to be assessed within the context of quality of care and have stated so in our submission. Results of the measure alone do not necessarily reflect the quality of care provided by hospitals but simply whether the total episode payments are greater than or less than would be expected for an average hospital with a similar case mix. Accordingly, measure scores are reported together with a quality signal (in this case mortality for the same condition) as an indication of the value of care. (CMS' mortality measures for these conditions were recently re-endorsed by NQF in the Fall 2020 cycle.) An example for one hospital is shown below; this hospital has payments that are greater than the national average and quality that is worse than the national rate, suggesting low-value care.

In addition, each spring, hospitals receive a detailed report of all the patients included in the measure, along with detailed breakdowns of post-acute care costs. Therefore, the payment

measures provide an opportunity for hospitals to explore the drivers of costs for their patients and assess the payment measure results in the context of the quality of care they provide to patients.

Proposed Committee Response:

Thank you for your comments. The Standing Committee and NQF recognize that cost and resource use measures should be used in the context of and reported with quality measures. The Standing Committee discussed the relationship between cost and quality measures, emphasizing the importance of reporting performance to demonstrate improvements in cost while ensuring similar or higher levels of care quality. Additionally, the current NQF cost and efficiency endorsement criteria do not require specifications or testing of a paired quality measure.

Action Item:

The Standing Committee should review the comment and developer's response and be prepared to discuss them in relation to the measure and its current recommendation status. The Standing Committee should determine whether they agree with the proposed response.

NQF Member Expression of Support

Throughout the 16-week continuous public commenting period, NQF members have the opportunity to express their support ('Support' or 'Do Not Support') for each measure to inform the Standing Committee's recommendations during the commenting period. This expression of support (or not) during the commenting period replaces the member voting opportunity that was previously held after Standing Committee deliberations. One NQF member provided their expressions of nonsupport (Appendix A.)

Appendix A: NQF Member Expression of Support Results

One NQF member provided their expressions of nonsupport for three of the five measures under consideration. Results for each measure are provided below.

#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/Yale CORE)

Member Council	Support	Do Not Support	Total
Health Professional	0	1	1

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

Member Council	Support	Do Not Support	Total
Health Professional	0	1	1

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

Member Council	Support	Do Not Support	Total
Health Professional	0	1	1