NATIONAL QUALITY FORUM
Driving measurable health improvements together

Cost and Efficiency, Spring 2021: Public and Member Comments

Funded by the Centers for Medicare and Medicaid Services under contract HHSM-500-2017-00060I Task Order 75FCMC19F0007

https://www.qualityforum.org
1099 14th Street NW, Suite 500 | Washington, DC 20005 | M 202.783.1300 F 202.783.3434

Contents

Cost and Efficiency, Spring 2021: Public and Member Comments 1
Measure-Specific Comments on Cost and Efficiency Spring 2021 Submissions 3
NQF #2431 Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction (AMI), Comment #1 3
Comment 3
Developer Response 4
NQF Response: 5
NQF Committee Response: 6
NQF #2436 Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for heart failure (HF), Comment #2 6
Comment 6
Developer Response: 7
NQF Response: 9
NQF #2579 Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia (PN), Comment #3 10
Comment 10
Developer Response 11
NQF Response: 13

Measure-Specific Comments on Cost and Efficiency Spring 2021 Submissions

NQF #2431 Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction (AMI), Comment #1

Standing Committee Recommendation: Recommended for Endorsement

Comment ID#: 7803

Commenter: Koryn Rubin, American Medical Association (AMA)

Council / Public: Health Professionals Council

Comment Period: Post-Evaluation Commenting Period

Date Comment was Submitted: 9/24/2021

Developer Response Required? Yes ☒ No ☐

Level of Support: Member Does Not Support

Themes: Reliability/Minimum Reliability Thresholds, Social Risk and Risk Adjustment, Cost and Quality Correlation

Comment

The American Medical Association (AMA) is concerned that this measure does not meet the scientific acceptability criteria and asks that the Standing Committee reconsider the current recommendation to continue endorsement of this measure. Specifically, the testing results, particularly for measure score reliability, empirical validity and the risk adjustment approach, do not provide the information needed to ensure that the measure produces the desired results.

Regarding the measure score reliability, we are concerned that the signal-to-noise value ranged from 0.298 to 0.594 and we do not support the current threshold of 0.4 used by CMS. The AMA believes that the minimum acceptable threshold should be 0.7 and the measure as specified does not meet this expectation.

The AMA strongly supports the tenet that cost must be assessed within the context of the quality of care provided; yet, the developer was unable to demonstrate that this measure correlates to any one quality measure within the hospital quality programs. This is particularly concerning since the submission clearly states that cost should not be evaluated alone and specific references are made to quality measures (e.g., mortality, readmissions) on which comparisons could be made. We are very troubled that this testing was not provided and we do not believe that cost measures against which no quality measure can be empirically assessed should achieve endorsement.

The AMA does not believe that the current risk adjustment model is adequate due to R-squared result of 0.078 nor is the measure adequately tested and adjusted for social risk factors. It is unclear to us why the developer would test social risk factors after adjusting for clinical risk factors rather than assessing the impact of both clinical and social risk factors in the model at the same time. These variations in how risk adjustment factors are examined could also impact how each variable (clinical or social) perform in the model and remain unanswered questions.

The AMA requests that these gaps in testing be addressed prior to endorsement of this measure. We appreciate the Committee's consideration of our comments.

Developer Response

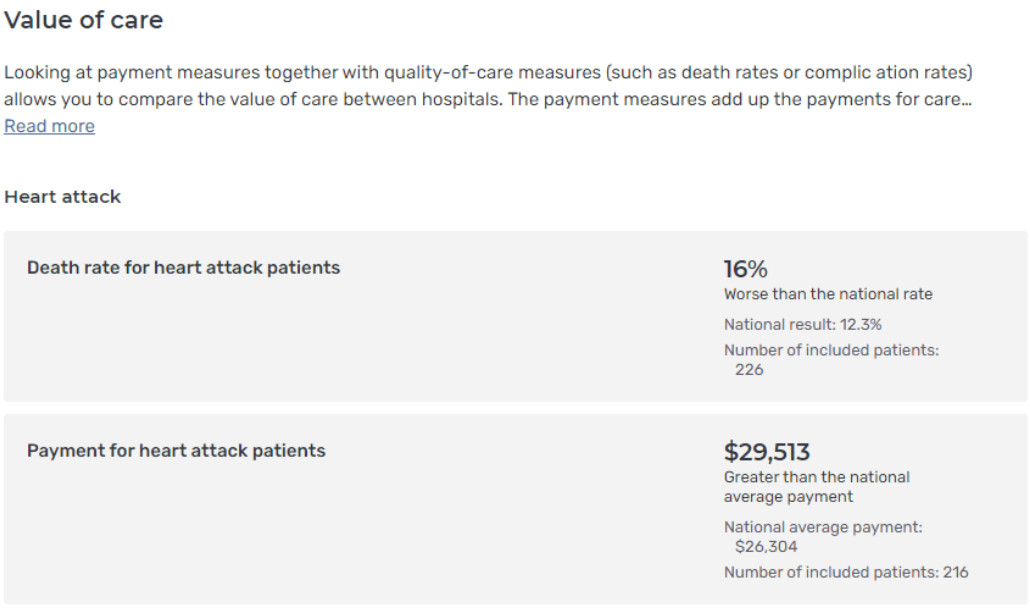
CORE's NQF submission fully satisfies NQF criteria for endorsement. NQF's Scientific Methods Panel (SMP) determined the measure is scientifically sound and passed the measure on both validity and reliability. The Cost and Efficiency Standing Committee voted in favor of re-endorsement of this measure. The developer notes that each of these issues were already addressed by the Standing Committee during the measure's review.

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).

References:

Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin, 1979, 86, 420-3428.

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.

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 Response:
Not Applicable

NQF 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 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.

The Standing Committee also 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's decision to pass the measure on validity and their input on the validity testing, and the approach to the risk adjustment modeling and agreed to recommend these measures for endorsement.

Lastly, 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.

NQF #2436 Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for heart failure (HF), Comment #2
Standing Committee Recommendation: Recommended for Endorsement

Comment ID#: 7804

Commenter: Koryn Rubin, American Medical Association (AMA)

Council / Public: Health Professionals Council

Comment Period: Post-Evaluation Commenting Period

Date Comment was Submitted: 9/24/2021

Developer Response Required? Yes ☒ No ☐

Level of Support: Member Does Not Support

Themes: Reliability/Minimum Reliability Thresholds, Social Risk and Risk Adjustment, Cost and Quality Correlation

Comment
The American Medical Association (AMA) is concerned that this measure does not meet the scientific acceptability criteria and asks that the Standing Committee reconsider the current recommendation to continue endorsement of this measure. Specifically, the testing results, particularly for measure score reliability, empirical validity and the risk adjustment approach, do not provide the information needed to ensure that the measure produces the desired results.

Regarding the measure score reliability, we are concerned that the signal-to-noise value ranged from 0.528 & 0.801 and we do not support the current threshold of 0.4 used by CMS. The AMA believes that the minimum acceptable threshold should be 0.7 and the measure as specified does not meet this expectation.

The AMA strongly supports the tenet that cost must be assessed within the context of the quality of care provided; yet, the developer was unable to demonstrate that this measure correlates to any one quality measure within the hospital quality programs. This is particularly concerning since the submission clearly states that cost should not be evaluated alone and specific references are made to quality measures (e.g., mortality, readmissions) on which comparisons could be made. We are very troubled that this testing was not provided and we do not believe that cost measures against which no quality measure can be empirically assessed should achieve endorsement.

The AMA does not believe that the current risk adjustment model is adequate due to R-squared result of 0.031 nor is the measure adequately tested and adjusted for social risk factors. It is unclear to us why the developer would test social risk factors after adjusting for clinical risk factors rather than assessing the impact of both clinical and social risk factors in the model at the same time. These variations in how risk adjustment factors are examined could also impact how each variable (clinical or social) perform in the model and remain unanswered questions.

The AMA requests that these gaps in testing be addressed prior to endorsement of this measure. We appreciate the Committee's consideration of our comments.

Developer Response:
CORE's NQF submission fully satisfies NQF criteria for endorsement. NQF's Scientific Methods Panel (SMP) determined the measure is scientifically sound and passed the measure on both validity and reliability. The Cost and Efficiency Standing Committee voted in favor of re-endorsement of this measure (14 voted yes, out of 15 members). The developer notes that each of these issues were already addressed by the Standing Committee during the measure's review.

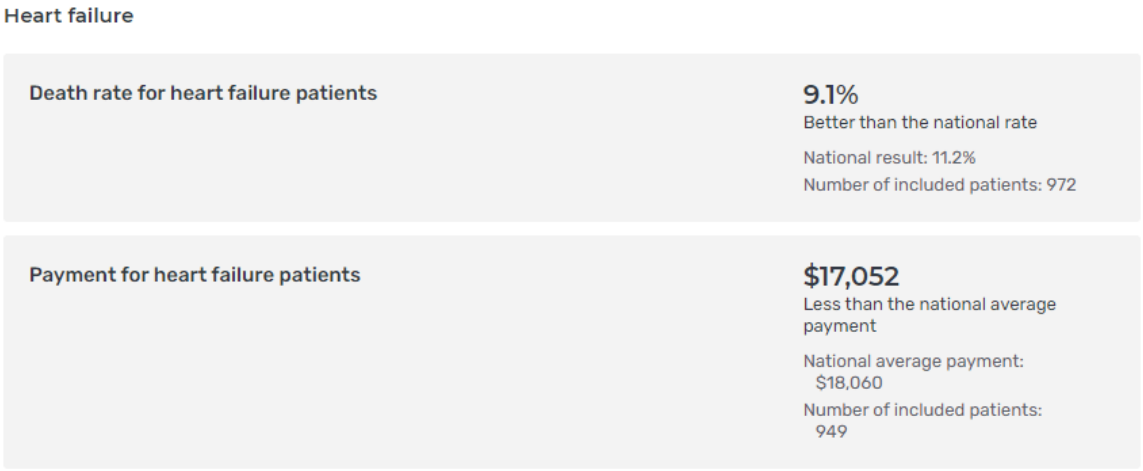
In our testing attachment, we provided split-sample reliability. To calculate split-sample reliability, we randomly sampled half of patients within each hospital from a three-year measurement period, calculated the measure for each hospital, and repeated the calculation using the second half of patients. 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).

References:

Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin, 1979, 86, 420-3428.

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.

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 Response:
Not Applicable**.**

NQF 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 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.

The Standing Committee also 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's decision to pass the measure on validity and their input on the validity testing, and the approach to the risk adjustment modeling and agreed to recommend these measures for endorsement.

Lastly, 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.

NQF #2579 Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia (PN), Comment #3

Standing Committee Recommendation: Recommended for Endorsement

Comment ID#: 7805

Commenter: Koryn Rubin, American Medical Association (AMA)

Council / Public: Health Professionals Council

Comment Period: Post-Evaluation Commenting Period

Date Comment was Submitted: 9/24/2021

Developer Response Required? Yes ☒ No ☐

Level of Support: Member Does Not Support

Themes: Reliability/Minimum Reliability Thresholds, Social Risk and Risk Adjustment, Cost and Quality Correlation

Comment

The American Medical Association (AMA) is concerned that this measure does not meet the scientific acceptability criteria and asks that the Standing Committee reconsider the current recommendation to continue endorsement of this measure. Specifically, the testing results, particularly for empirical validity and the risk adjustment approach, do not provide the information needed to ensure that the measure produces the desired results.

The AMA does not believe that the current risk adjustment model is adequate due to R-squared result of 0.076 nor is the measure adequately tested and adjusted for social risk factors. It is unclear to us why the developer would test social risk factors after adjusting for clinical risk factors rather than assessing the impact of both clinical and social risk factors in the model at the same time. These variations in how risk adjustment factors are examined could also impact how each variable (clinical or social) perform in the model and remain unanswered questions.

The AMA strongly supports the tenet that cost must be assessed within the context of the quality of care provided; yet, the developer was unable to demonstrate that this measure correlates to any one quality measure within the hospital quality programs. This is particularly concerning since the submission clearly states that cost should not be evaluated alone and specific references are made to quality measures (e.g., mortality, readmissions) on which comparisons could be made. We are very troubled that this testing was not provided and we do not believe that cost measures against which no quality measure can be empirically assessed should achieve endorsement.

The AMA requests that these gaps in testing be addressed prior to endorsement of this measure. We appreciate the Committee's consideration of our comments.

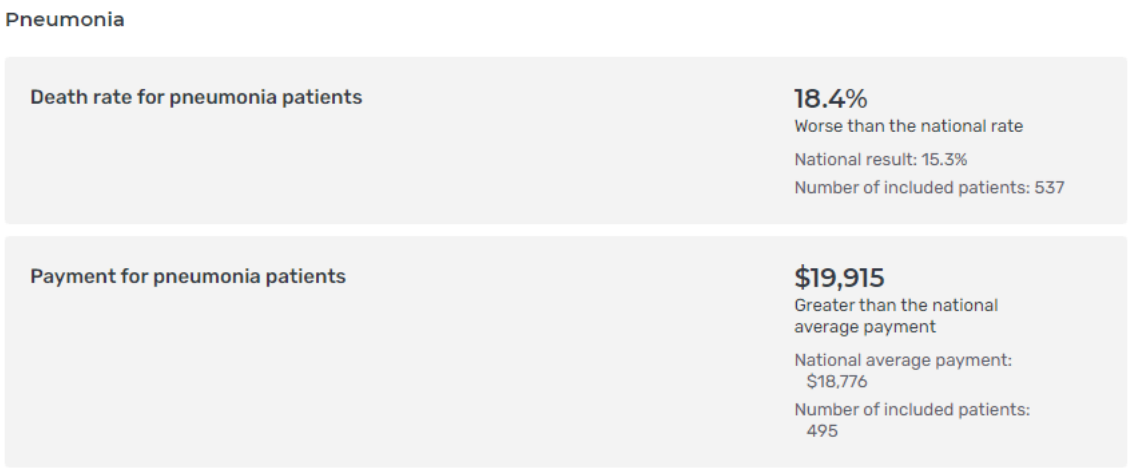
Developer Response

CORE's NQF submission fully satisfies NQF criteria for endorsement. NQF's Scientific Methods Panel (SMP) determined the measure is scientifically sound and passed the measure on both validity and reliability. The Cost and Efficiency Standing Committee voted in favor of re-endorsement of this measure (15 voted yes, out of 15 members). The developer notes that each of these issues were already addressed by the Standing Committee during the measure's review.

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 the pneumonia measure, the split-sample reliability was 0.815, which falls within the thresholds currently under consideration by the Scientific Methods Panel (SMP).

References: **Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin, 1979, 86, 420-3428.**

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.

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.

NQF Response:
Not Applicable.

NQF 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 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.

The Standing Committee also 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's decision to pass the measure on validity and their input on the validity testing, and the approach to the risk adjustment modeling and agreed to recommend these measures for endorsement.

Lastly, 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.