NATIONAL QUALITY FORUM

Memo

September 23, 2019

To: Surgery Standing Committee

From: NQF staff

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

Purpose of the Call

The Surgery Standing Committee will meet via web meeting on September 23, 2019 from 1:00 pm – 3:00 pm ET. The purpose of this 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' expressions of support of the measures under consideration

Standing Committee Actions

1. Review this briefing memo and draft report.
2. Review and consider the full text of all comments received.
3. Review the NQF members' expressions of support of the submitted measures.
4. Be prepared to 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:

Speaker dial-in #: (800) 768-2983; **Access code**: 7445915
Weblink: https://core.callinfo.com/callme/?ap=8007682983&ac=7445915&role=p&mode=ad

Background

The report reflects the review of measures in the surgery project. The measures in NQF's surgery endorsement project focus on key surgical care processes across an array of procedure types that include outcomes for general and subspecialty procedures, including cardiac, orthopedic, ophthalmological, and vascular surgeries and procedures, and all phases of perioperative care. In this project, measures focused on lung resection and lobectomy for lung cancer and hospital visits following general surgery procedures.

The 22-member Surgery Standing Committee oversees the NQF Surgery measure portfolio, evaluating both newly submitted and previously endorsed measures against NQF's measure evaluation criteria, identifying gaps in the measurement portfolio, and providing feedback on how the portfolio should evolve.

On July 5, 10, and 15, 2019, the Surgery Standing Committee evaluated five maintenance measures and two new measures against NQF's criteria. The Committee recommended all seven measures for endorsement.

Comments Received

NQF solicits comments on measures undergoing review in various ways and at various times throughout the evaluation process. First, NQF solicits comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). Second, NQF solicits member and public comments during a 16-week comment period via an online tool on the project webpage.

Pre-evaluation Comments

NQF solicits comments prior to the evaluation of the measures via an online tool on the project webpage. For this evaluation cycle, the pre-evaluation comment period was open from May 8, 2019 to June 19, 2019 for the measures under review. All of these pre-evaluation comments were provided to the Committee prior to the measure evaluation meeting.

Post-evaluation Comments

The draft report was posted on the project webpage for public and NQF member comment on August 13, 2019 for 30 calendar days. During this commenting period, NQF received two comments from the American Medical Association (AMA) (Appendix B). The developer's responses to the comments received are included as Appendix C.

Summary of Comments

3493 Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) for Merit-Based Incentive Payment System (MIPS) Eligible Clinicians and Eligible Clinician Groups

The AMA does not support endorsement of the measure and has concerns that the measure does not meet the evidence and scientific acceptability criteria. Specifically, the AMA commented that:

- "Insufficient evidence was provided to support attribution of the measure to physicians or practices;
- The measure score reliability results are too low when based on the minimum case number of 25 patients. Measures should meet minimum acceptable thresholds of 0.7 for reliability;
- The conceptual basis used to explain which social risk factors were tested in Section 2b3.3a is inadequate and additional testing is needed to evaluate clinical factors in conjunction with social risk factors as well as the impact that the inclusion of these factors had on the absolute change of the rates; and
- Additional testing is needed to demonstrate how the measure would perform under the MIPS benchmark methodology and Physician Compare Star Ratings since CMS utilizes two different methodologies for ranking and profiling physicians."

Measure Steward/Developer Response:
The developer thanks AMA for the comment and has addressed each of the concerns separately. The developer's complete responses have been included as Appendix C.

Proposed Committee Response:
Thank you for your comment. The Committee will discuss the comment in its deliberations during the September 23 post-comment web meeting.

3494 Hospital 90-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Coronary Artery Bypass Graft (CABG) Surgery

The AMA does not support endorsement of the measure and has concerns that the measure does not meet the scientific acceptability and usability and use criteria. Specifically, the AMA commented that:

- "The measure score reliability results are too low when based on the minimum case number of 25 admissions. Measures should meet minimum acceptable thresholds of 0.7 for reliability;
- The conceptual basis used to explain which social risk factors were tested in Section 2b3.3a is inadequate and additional testing is needed to evaluate clinical factors in conjunction with social risk factors as well as the impact that the inclusion of these factors had on the absolute change of the rates; and
- It remains unclear whether a measure that currently only identifies small differences in performance scores enables users to distinguish meaningful differences in performance. Specifically, the 10th percentile yields a rate 1.08% lower and 90th percentile is 1.74% higher than an average facility with a similar patient mix."

Measure Steward/Developer Response:
The developer thanks AMA for the comment and has addressed each of the concerns separately. The developer's complete responses have been included as Appendix C.

Proposed Committee Response:
Thank you for your comment. The Committee will discuss the comment in its deliberations during the September 23 post-comment web meeting.

NQF Member Expression of Support

Throughout the 16-week continuous public commenting period, NQF members had the opportunity to express their support ('support' or 'do not support') for each measure submitted for endorsement consideration to inform the Committee's recommendations. One NQF members provided their expressions of support: See Appendix A.

Appendix A: NQF Member Expression of Support Results

One NQF member provided an expression of nonsupport of two measures. Results for each measure are provided below.

3493 Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) for Merit-Based Incentive Payment System (MIPS) Eligible Clinicians and Eligible Clinician Groups (Centers for Medicare and Medicaid Services)

Member Council	Support	Do Not Support	Total
Consumer			
Health Plan			
Health Professional		1	
Provider Organization			
Public/Community Health Agency			
Purchaser			
QMRI			
Supplier/Industry			

3494 Hospital 90-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Coronary Artery Bypass Graft (CABG) Surgery

Member Council	Support	Do Not Support	Total
Consumer			
Health Plan			
Health Professional		1	
Provider Organization			
Public/Community Health Agency			
Purchaser			
QMRI			
Supplier/Industry			

Appendix B: Comments Received

Measure-Specific Comments

3493 Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) for Merit-Based Incentive Payment System (MIPS) Eligible Clinicians and Eligible Clinician Groups

Vote: Do not support

The American Medication Association (AMA) appreciates the Standing Committee discussion and evaluation of this measure but continues to have significant concerns regarding whether the measure meets the NQF Measure Evaluation Criteria, particularly for evidence and scientific acceptability.

As mentioned in our comments submitted prior to the committee's evaluation, we believe that:

- Insufficient evidence was provided to support attribution of the measure to physicians or practices;
- The measure score reliability results are too low when based on the minimum case number of 25 patients. Measures should meet minimum acceptable thresholds of 0.7 for reliability;
- The conceptual basis used to explain which social risk factors were tested in Section 2b3.3a is inadequate and additional testing is needed to evaluate clinical factors in conjunction with social risk factors as well as the impact that the inclusion of these factors had on the absolute change of the rates; and
- Additional testing is needed to demonstrates how the measure would perform under the MIPS benchmark methodology and Physician Compare Star Ratings since CMS utilizes two different methodologies for ranking and profiling physicians.

As a result, the AMA is unable to support endorsement of the measure at this time and requests that the Committee reconsiders its recommendation for endorsement.

3494 Hospital 90-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Coronary Artery Bypass Graft (CABG) Surgery

Vote: Do not support

The American Medication Association (AMA) appreciates the Standing Committee discussion and evaluation of this measure but continues to have significant concerns regarding whether the measure meets the NQF Measure Evaluation Criteria, particularly for scientific acceptability and usability and use.

As mentioned in our comments submitted prior to the committee's evaluation, we believe that:

- The measure score reliability results are too low when based on the minimum case number of 25 admissions. Measures should meet minimum acceptable thresholds of 0.7 for reliability;

- The conceptual basis used to explain which social risk factors were tested in Section 2b3.3a is inadequate and additional testing is needed to evaluate clinical factors in conjunction with social risk factors as well as the impact that the inclusion of these factors had on the absolute change of the rates; and
- It remains unclear whether a measure that currently only identifies small differences in performance scores enables users to distinguish meaningful differences in performance. Specifically, the 10^{th} percentile yields a rate 1.08% lower and 90^{th} percentile is 1.74% higher than an average facility with a similar patient mix.

As a result, the AMA is unable to support endorsement of the measure at this time and requests that the Committee reconsiders its recommendation for endorsement.

Appendix C: Measure Steward/Developer Response

Measure-Specific Responses

3493 Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) for Merit-Based Incentive Payment System (MIPS) Eligible Clinicians and Eligible Clinician Groups

We appreciate your comments and have addressed each of your concerns below, separately.

Attribution

We also agree with the conclusions outlined within NQF's final report, Improving Attribution Models (NQF, 2018), in that attribution models should reflect clinicians and providers with reasonable influence on the care and outcomes for patients in order to enforce accountability and facilitate quality improvement. During development, we solicited a wide variety of clinician, technical, and patient feedback through stakeholder engagement. The Technical Expert Panel, in particular, felt strongly that it was appropriate to attribute readmissions to the billing surgeon to encourage coordination and shared accountability.

Reliability Testing

We agree that it is important that the final volume threshold correspond to adequate reliability. Constructing meaningful, reliable, valid provider quality measures is challenging and requires balancing competing factors and values. In section 2a2.3 of the NQF submission form we report: "Entity-level reliability testing indicated that for entities with at least 25 procedures, the median signal-to-noise ratio reliability was 0.793 [IQR 0.695 – 0.878] for clinicians and 0.790 [IQR 0.647 – 0.907] for clinician groups." The ranges, not reported here, are [0.582 – 0.988] and [0.463 – 0.996]. According to Landis and Koch (1977) reliability of 0.4 or more is 'fair'. Thus, even for the least reliable values the 25-volume threshold provides fair reliability We believe this is evidence that these measures do capture reliable quality signals at the clinician and group level under the proposed attribution.

Validity Testing

We included incorrect information in the face validity section of the submission form, and apologize for the confusion. The measure is fully specified and the measure development process is complete, and the actual survey results differ from those reported. We conducted face validity on the Final Attribution Rule and on the MIPS Eligible Clinician and Eligible Clinician Group Measure Scores. The Technical Expert Panel (TEP) strongly supported attribution to the Billing Surgeon. All 19 TEP members asked to complete a survey regarding validity and usability of the MIPS HKC measure, 16 responded; their responses are reported in the following table.

Table 1. TEP reports of agreements

The HKC:	Strongly Disagree	Moderately Disagree	Somewhat Disagree	Somewhat Agree	Moderately Agree	Strongly Agree
…measure scores are valid and useful	1	0	2	3	9	1
…measure will provide info to be used for quality improvement	1	1	2	4	5	3

As shown in Table 1, the majority of the respondents, 13/16 or 81%, agreed that the HKC measure scores were valid and useful, and 12/16 or 75% agreed that the measure would provide information that could be used to improve the quality of care.

Among those who disagreed, the primary concern was that the lowest volume eligible clinicians would not be measured, rather than concern with the measure itself. Though this is a challenge with all quality measures, it may be of particular concern when there may be an inverse relationship between volume and quality. It is notable that even with the 25-patient volume threshold, over 96% of patients are retained; it is also important to note that the measure counts only Medicare Fee-For-Service patients, so the total case volume of those eligible clinicians excluded by the volume threshold is unknown, and could be quite high.

Overall, the survey indicates support of the validity and usability of the measure.

Again, the measure is fully specified and the measure development process is complete. We apologize for the typo error and have requested the removal of the sentence in question, the last sentence of Section 2b1.4.

Social Risk Factor Testing

We tested for the effects of including two social risk factors within the model (dual eligibility status and low Agency for Healthcare Research and Quality SES) on final risk-adjusted rates for clinicians and clinician groups. The correlation between the adjusted and unadjusted scores for clinicians and clinician groups were 0.99, indicating extremely high agreement and that adding these social risk factors would have minimal impact on measure scores. Ongoing research aims to identify valid patient-level social risk factors and highlight disparities related to social risk. As additional variables become available, they will be considered for testing and inclusion within the measure.

Since the release of the Evaluation of the NQF Trial period for Risk Adjustment for Social Risk Factors report in July 2017, NQF announced the launch of a new, three-year initiative to explore unresolved issues that surfaced in the 2015-2017 social risk factor trial.[a] The stated goal of the new Social Risk Trial is to "help inform a decision on whether to permanently change NQF's policy to allow social risk adjustment for outcome measures."[b] For risk-adjusted outcome measures, CMS first considers adjustment for clinical conditions and then examines additional risk imparted by social risk factors after the potential for greater disease burden is included in the risk model. We believe that this is consistent with NQF current guidance and is appropriate

---

[a] National Quality Forum (NQF). NQF Statement on Board of Directors Decision Regarding Social Risk Trial, http://www.qualityforum.org/News_And_Resources/Press_Releases/2017/NQF_Statement_on_Board_of_Directors_Decision_Regarding_Social_Risk_Trial.aspx

[b] National Quality Forum (NQF). Social Risk Trial FAQ, June 28, 2018. http://www.qualityforum.org/ProjectMaterials.aspx?projectID=87820. Accessed September 9, 2019.

given the evidence cited in our submission that people who experience greater social risk are more likely to have more disease burden compared with those who do not; and that this is clearly not a signal of hospital quality. In addition, according to NQF guidance, developers should assess social risk factors for their contribution of unique variation in the outcome – that they are not redundant.[c] Therefore, if clinical risk factors explain all or most of the patient variation in the outcome, then NQF guidance does not support adding social risk factors that do not account for variation.

In addition to the correlation between adjusted and unadjusted scores, we also tested the change in risk-adjusted readmission rates. When incorporating the duel eligible risk factor, risk-adjusted readmission rates dropped an absolute value of --0.0046% for clinicians and -0.0039% for clinician groups. When incorporating low AHRQ SES, risk-adjusted readmission rates dropped an absolute value of -0.0022% for clinicians and -0.0023% for clinician groups.

Program-Specific Testing
NQF doesn't specify or require testing for impact on program inclusion, program benchmarking, or star rating systems. At this time, it is not known how CMS will use this measure in the MIPS program.

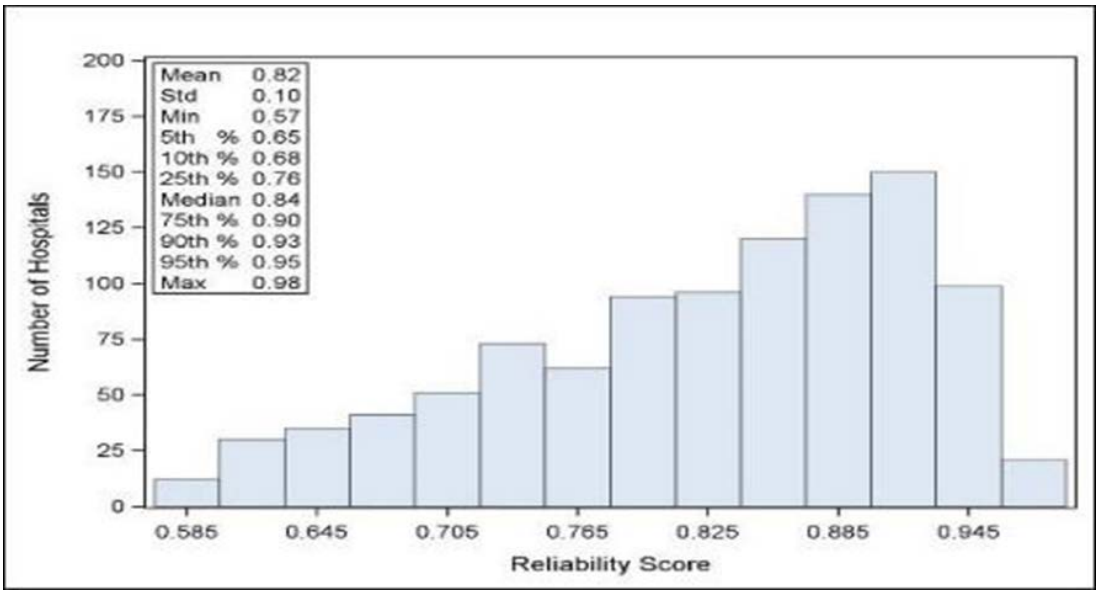
Conclusion
We agree with the importance of balancing these competing considerations. We are committed to constant refinement and improvement of risk adjustment models used in all measures. We will reevaluate this model and available risk factors on an ongoing basis, with the goal of producing the most accurate and fair risk adjustment models for assessing provider performance.

3494 Hospital 90-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Coronary Artery Bypass Graft (CABG) Surgery
Response:

We appreciate your comments and have addressed each of your concerns below.
We agree that it is important that the final volume threshold correspond to adequate reliability. In section 2a2.3 of the NQF submission form we report: "the signal to noise reliability score for each hospital with at least 25 admissions (see Figure 1 below). The median reliability score was 0.84, ranging from 0.57 to 0.98. The 25th and 75th percentiles were 0.76 and 0.90, respectively. The median reliability score demonstrates high reliability between the two samples." We believe this is evidence that the measure does capture reliable quality signals with 25 or more cases.

[c] National Quality Forum (NQF). Risk adjustment for socioeconomic status or other sociodemographic factors: Technical report. 2014; http://www.qualityforum.org/Publications/2014/08/Risk_Adjustment_for_Socioeconomic_Status_or_Other_Sociodemographic_Factors.aspx. Accessed September 3, 2019.

Figure 1. Measure Score Reliability Testing: Signal to Noise Ratio Results



Literature indicates that the relationship between patient social risk factors and mortality are multifaceted and causal pathways include patient health upon admission, social risk factors outside of the hospital, quality of hospitals, and differential care within a hospital. Overall quality of hospitals and differential care within a hospital should be captured by mortality outcome measures. Health status upon admission is accounted for through exclusions for patient severity beyond the influence of quality care at a hospital and the risk model, applying adjustment for variables clinically relevant to mortality following coronary artery bypass graft (CABG). To address the potential impact of social risk factors outside of the hospital, we tested for the effects of including two social risk factors within the model (dual eligibility status and low Agency for Healthcare Research and Quality [AHRQ] SES) on final risk-adjusted rates for hospitals. We found no significant impact of either of these indicators on model performance and their addition is unlikely to affect hospital profiling.

CMS and NQF have previously reviewed literature and conducted research to identify available and reliable social risk factor variables. Few options were found to be reliably collected and representative of a patient's specific socioeconomic status, rather than their race or ethnicity. While the available social risk factors are limited, we believe the variables tested cover both patients' environment and specific situations. The AHRQ SES index is derived from census block group level data and linked to patient zip codes to capture environmental and community factors. Dual eligibility status identifies patients that quality for both Medicare (indicating 65 years of age or older) and Medicaid (indicating low SES or disability).

Ongoing research aims to identify valid patient-level social risk factors and highlight disparities related to social risk. As additional variables become available, they will be considered for testing and inclusion within the measure.

Since the release of the Evaluation of the NQF Trial period for Risk Adjustment for Social Risk Factors report in July 2017, NQF announced the launch of a new, three-year initiative to explore

unresolved issues that surfaced in the 2015-2017 social risk factor trial.[d] The stated goal of the new Social Risk Trial is to "help inform a decision on whether to permanently change NQF's policy to allow social risk adjustment for outcome measures."[e] For risk-adjusted outcome measures, CMS first considers adjustment for clinical conditions and then examines additional risk imparted by social risk factors after the potential for greater disease burden is included in the risk model. We believe that this is consistent with NQF current guidance and is appropriate given the evidence cited in our submission that people who experience greater social risk are more likely to have more disease burden compared with those who do not; and that this is clearly not a signal of hospital quality. In addition, according to NQF guidance, developers should assess social risk factors for their contribution of unique variation in the outcome – that they are not redundant.[f] Therefore, if clinical risk factors explain all or most of the patient variation in the outcome, then NQF guidance does not support adding social risk factors that do not account for variation.

We agree with the importance of balancing these competing considerations. We are committed to constant refinement and improvement of risk adjustment models used in all measures. We will reevaluate this model and available risk factors on an ongoing basis, with the goal of producing the most accurate and fair risk adjustment models for assessing provider performance.

Mortality is an important health outcome that is meaningful to patients and providers. The median hospital-level risk standardized mortality rate (RSMR) is 4.67%, meaning 4.67% or more patients are expected to die within 90 days following CABG procedure. The hospital-level variation in performance on the measure score between the lowest (RSMR of 2.04) and highest (RMSR of 11.26) performing hospitals shows there is a meaningful difference across hospitals in 90-day all-cause mortality following CABG procedure, a clear quality gap. Furthermore, the median odds ratio suggests a meaningful increase in risk of death if the procedure was performed at a lower performance hospital compare to a higher performance hospital. A patient has a 47% increase in odds of death following CABG procedure at a lower performance hospital than a higher performance hospital, which indicates the impact of quality on the outcome rate is substantial.

---

[d] National Quality Forum (NQF). NQF Statement on Board of Directors Decision Regarding Social Risk Trial, http://www.qualityforum.org/News_And_Resources/Press_Releases/2017/NQF_Statement_on_Board_of_Directors_Decision_Regarding_Social_Risk_Trial.aspx

[e] National Quality Forum (NQF). Social Risk Trial FAQ, June 28, 2018. http://www.qualityforum.org/ProjectMaterials.aspx?projectID=87820. Accessed September 9, 2019.

[f] National Quality Forum (NQF). Risk adjustment for socioeconomic status or other sociodemographic factors: Technical report. 2014; http://www.qualityforum.org/Publications/2014/08/Risk_Adjustment_for_Socioeconomic_Status_or_Other_Sociodemographic_Factors.aspx. Accessed September 3, 2019.