NATIONAL QUALITY FORUM

Memo

May 3, 2019

To: Surgery Standing Committee

From: NQF staff

Re: Post-comment web meeting to re-vote on "Consensus Not Reached" (CNR) Measures

Purpose of the Call

The Surgery Standing Committee will meet via web meeting on May 8, 2019 from 1:00 pm to 3:00 pm ET. The purpose of this call is to:

- Review comments received during the post-evaluation public and member comment period and provide input on proposed responses;
- Re-vote on "Consensus Not Reached" measures;
- Review and discuss related and competing measures;
- Determine whether reconsideration of any measures or other courses of action are warranted.

Standing Committee Actions

1. Review this briefing memo and the draft report.
2. Review and consider the full text of comments received and provide feedback on the proposed responses to the post-evaluation comments.

Conference Call Information

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

Speaker dial-in #: 1-800-768-2983
Access code: 2511568
Web link: https://core.callinfo.com/callme/?ap=8007682983&ac=2511568&role=p&mode=ad

Background

On February 13 and February 20, 2019, the Surgery Standing Committee evaluated 15 measures undergoing maintenance. The Committee recommended 13 measures for endorsement and did not reach consensus on two measures.

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 December 11, 2018 to January 30, 2019 for the measures under review. No pre-evaluation comments were submitted 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 March 21, 2019 for 30 calendar days. During this commenting period, NQF received one comment from a member organization:

Member Council	# of Member Organizations Who Commented
Consumer	0
Health Plan	0
Health Professional	1
Provider Organization	0
Public/Community Health Agency	0
Purchaser	0
QMRI	0
Supplier/Industry	0

We have included the post-evaluation comments we received in the excel spreadsheet posted to the Committee SharePoint site. The spreadsheet contains the commenter's name, comment, and draft response for the Committee's consideration. Please review this table before the meeting and consider the individual comments received and the proposed responses.

Comments and their Disposition

General Comment

The Society of Thoracic Surgeons (STS) appreciates the opportunity to comment on the overarching issues described in the Surgery Standing Committee report on its recent evaluation of 15 STS measures.

Levels of analysis: The meeting summary report states the following: "The developer confirmed that physicians are the accountable entity for these measures rather than hospital/facilities. However, NQF guidance states that the level of analysis must align with testing; therefore, 'hospital/facilities' will be removed from the specifications. Additional testing at the facility level is required for endorsement at both levels of analysis." This statement is inaccurate. None of these measures were designed for individual physicians, but rather for physician group practices and—at the option of these practices—the facilities/hospitals at which they perform surgery. That point was made clear by all STS representatives at the meeting, who have been intimately involved in the development of these measures.

Race and risk adjustment: As noted at the Committee meetings in February, the STS contends that it remains appropriate to include race in our risk models, not as a sociodemographic factor (nor as a surrogate for such factors), but as one of various preoperative variables that are independently and significantly associated with clinical outcomes. Race has an empirical association with outcomes and has the potential to confound the interpretation of a hospital's outcomes, although the underlying mechanism is unknown (e.g., genetic factors, differential effectiveness of certain medications, rates of certain associated diseases not accounted for in the risk models, and racial differences in vessel anatomy and suitability for bypass). This is similar to the well-known fact that female gender is associated with worse outcomes and is included in our CABG models (e.g., their coronary arteries tend to be smaller and more challenging for anastomoses). For future submissions, a reasonable compromise would be to present results with adjustment for race as well as results stratified by race but without race adjustment.

Score- level validity testing methodology: The meeting summary states that "...star-rating consistency over time is expected and is not an appropriate approach to demonstrating validity." Our major validity indicator is the association of our 1, 2, and 3 star (worse than expected, as expected, and better than expected) composite ratings with the relevant mortality and morbidity scores, which we regard as the "gold standard."

Public reporting and transparency: With all of our outcome measures, the STS seeks to produce consistent, credible results that discriminate between significant differences in performance and facilitate informed decision-making, as required by NQF criteria. Data analysis for the first STS composite measure (1) demonstrated that risk-adjusted mortality, estimated separately, was able to statistically discriminate only 1% of providers as outliers, whereas the CABG composite (which also includes process measures and a morbidity domain) was able to discriminate 23%. A more recent analysis conducted for our newest publicly-reported composite (mitral repair/replacement) showed that, based on mortality data alone, the performance of less than 1% of surgical programs could reliably be classified as significantly higher or lower than the STS mean score; the mortality-morbidity composite classified 8.3% of programs as high or low performers (2). We have therefore concluded that it is more clinically meaningful to publicly report operative mortality in a composite with other quality metrics rather than reporting each item separately. The same reasoning applies to components of the composite morbidity domain, most of which have occurrence rates in the same range as that of mortality. If publicly reported as individual risk-adjusted measures, they would effectively be useless to patients in distinguishing quality differences among providers. The STS decision to not publicly report operative mortality alone or individual complication rates is not based solely on the statistical analyses described above. Qualitatively, the any-or-none approach to the morbidity composite domain is also a far more demanding and patient-centric standard. For patients and their families, it is much more relevant to know how best to avoid not just one or two of the major complications, but all of them. The composite therefore provides the likelihood that they will achieve this goal at different institutions. Reporting individual rates with inevitably wide confidence intervals would have greater probability of misleading rather than informing patients.

Patient and consumer perspective: The STS agrees that easy-to-access, meaningful information on provider performance is essential to enable patients to make informed decisions about their healthcare. It is for this reason that we continue to publicly report our composite measures as described above and are among the leaders in public reporting across all medical specialties. Additionally, following the Surgery Standing Committee meetings in February, we took immediate steps to expand definitions and other explanatory information on our public reporting web pages to enhance the transparency of composite results reported online. We also plan to expand the educational and quality-related information available on our patient website (The Patient Guide to Heart, Lung, and Esophageal Surgery) to assist patients with treatment options and decision-making related to cardiothoracic surgery.

1. Quality Measurement in Adult Cardiac Surgery: Part 2-Statistical Considerations in Composite Measure Scoring and Provider Rating. Brien SM, Shahian DM, DeLong ER, et al. (2007) Annals of Thoracic Surgery, 83 (4 SUPPL.), pp. S13-S26.

2. The Society of Thoracic Surgeons Mitral Repair/Replacement Composite Score: A Report of the Society of Thoracic Surgeons Quality Measurement Task Force. Badhwar V, Rankin JS, He X, et al. (2016) Annals of Thoracic Surgery, 101 (6), pp. 2265-2271.

Proposed NQF Response:

Level of analysis: NQF criteria requires that testing be provided for all the levels specified and intended for measure implementation (e.g., individual clinician, group/practice, hospital/facility, health plan, etc.). The developer conducted testing at the clinician group/practice level; therefore, the measures will be re-endorsed at this level of analysis. Testing was not conducted at the hospital/facility level; thus, the measures will not be endorsed at the hospital level of analysis.

Proposed NQF Response:

Race and risk adjustment: In 2014, NQF's Expert Panel on Risk Adjustment for Sociodemographic Factors determined that the effects of race and ethnicity are confounded by socioeconomic status (SES) and should not be used as proxies for SES (*Socioeconomic Status or Other Sociodemographic Factors* Technical Report, p. 42). The Expert Panel acknowledged that some see race and ethnicity like other potential confounders but recommended careful thought, consideration, and a clear rationale be used when adjusting performance measures for race and ethnicity because of concerns about bias and racism. The Expert Panel also encouraged reporting of data stratified by race and ethnicity to assess and address disparities in healthcare. If the developer provides stratified measure results for future submissions then stratification variables, definitions, specific data collection items/responses, etc. are required.

During the initial phase of the social risk trial, the Disparities Standing Committee provided additional guidance on the use of race and ethnicity as risk factors. Standing Committee members and members of the public raised concerns that some measures may have used race as proxy for socioeconomic status. Guidance from the Disparities Standing Committee stressed that race should not be used as a proxy for SES; however, there may be certain biological reasons when race could be an appropriate clinical

factor to include in a risk-adjustment model (e.g., potential tumor characteristics in African-American women with breast cancer).

As part of the social risk trial measure developers **are required** to provide a conceptual rationale describing the relationship between a social risk factor and the outcome of interest. If a conceptual relationship exists, developers should conduct empirical analyses to examine the relationship between the social risk factor and the outcome of interest.

Proposed NQF and Committee Response:

Score- level validity testing methodology: The NQF Scientific Methods Panel, made up of individuals with methodologic expertise, determined that star-rating consistency over time is not an appropriate approach to demonstrating validity and questioned the utility of the content validation approach used by the developer. The Methods Panel did not reach consensus on the validity of the measures. The Committee discussed the validity and determined that the results were acceptable. NQF and the Committee recommend that STS explore other types of analysis to strengthen the demonstration of validity for future submissions.

Proposed NQF Response:

Public Reporting and Transparency: Component measures in a composite measure are not required to be NQF-endorsed. NQF-endorsed measures are required to be used in at least one accountability application within three years after initial endorsement and publicly reported within six years after initial endorsement. This must-pass criterion (accountability and transparency) for maintenance measures is under advisement by the CSAC and may change in the future.

Proposed Committee Response:

Patient and consumer perspective: The Committee appreciates STS's efforts to improve the quality of their publicly available information so patients and their families, and other consumers can make more informed decisions about their healthcare. The Committee looks forward to working with STS to continue improving the quality of surgical care and publicly available data.

Committee Action Items:

The Committee must re-vote on Use (must-pass criterion for maintenance measures) for 0122 *Risk-Adjusted Operative Mortality for Mitral Valve (MV) Replacement + CABG Surgery* and 0114 *Risk-Adjusted Postoperative Renal Failure.* If the total PASS votes represent greater than 60 percent of a quorum, the measures pass the criterion, and the Committee must vote on Usability and vote on whether to recommend endorsement.

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. No expressions of support were received from the NQF members.