NATIONAL QUALITY FORUM
Driving measurable health improvements together

Meeting Summary

Scientific Methods Panel Measure Evaluation Web Meeting – Fall 2020

The National Quality Forum (NQF) convened the Scientific Methods Panel (SMP) on October 28-29, 2020 for a discussion of the scientific properties (reliability and validity) of several complex measures submitted to the fall 2020 evaluation cycle. Of the 25 measures reviewed by the SMP this cycle, eight measures were discussed during this meeting, including those for which subgroup[1] members did not reach consensus in their preliminary evaluations, those that did not initially pass the SMP evaluation but for which measure developers provided additional information, and those that were otherwise pulled for discussion by the SMP subgroup members. The SMP also discussed several overarching methodological issues which were identified based on their review of the measures this cycle. A brief summary of the discussion of these issues, the eight measures discussed during the web meeting, and voting results for the measures not discussed during the webinar are included in this document.

Welcome, Introductions, and Review of Meeting Objectives

Sai Ma, NQF managing director and senior technical expert, welcomed the members of the SMP and measure developers, other NQF staff, and members of the public to the web meeting. Sheri Winsper, NQF senior vice president of Quality Measurement, and SMP co-chairs David Nerenz and Christie Teigland, also provided opening remarks. Hannah Ingber, NQF senior analyst, asked SMP members to introduce themselves and provide any disclosures or conflict of interest relevant to the measures to be discussed during the meeting. Panel member Jack Needleman identified a conflict of interest with measures #0202 and #0141. Panel member Patrick Romano identified a conflict of interest with measure #0531. Panel member Zhenqiu Lin identified a conflict of interest with measure #0505. These Panel members did not participate via chat or verbally during discussion of the respective measures. Dr. Ma then described the process for the measure discussions and reviewed relevant NQF evaluation criteria.

Discussion of Overarching Methodological Issues Identified During Measure Evaluation

Acceptable Thresholds for Reliability

The members returned to the discussion of acceptable thresholds for reliability. During previous meetings, the SMP has not come to consensus on whether setting standardized thresholds for all reliability testing is appropriate. Members pointed out that the set of adjectives used to assess reliability as based on Landis and Koch's work *The Measurement of Observer Agreement for Categorical Data* (Biometrics. Vol. 33, No. 1 (1977), 159-174) is arbitrary. It may only be appropriate for inter-rater reliability (IRR) testing, not for measure score testing.

[1] Subgroups are determined by the NQF SMP team with input from the SMP co-chairs. NQF assigns measures to subgroup members for evaluation based on panelists' relevant expertise, interests, measure specific disclosures of interest, and CDP standing committee membership.

This conversation relates to another discussion SMP members have held in prior meetings. In previous meetings, SMP members have cited some concerns about evaluating measures without consideration of their use. For example, measures used to calculate penalties in payment programs may require stronger reliability, especially at the physician level. Measures that can only reliably identify extreme outliers cannot be used to reliably assign providers to groups like quintiles or deciles. The thresholds for reliability may need to vary based on use of the measure. However, the SMP does not evaluate the NQF criteria of use, and the NQF endorsement process is currently agnostic to future measure use. Therefore, setting thresholds while being agnostic to purpose may be difficult. One member suggested that the submission form asks developers explicitly for the standard error of measurement to show what amount of variance is being reliably explained. Further investigation into this issue is needed prior to any resolution. The SMP understands that any proposed reliability thresholds coming from the SMP in the future would require Consensus Standards Approval Committee (CSAC) and the Board's approval before becoming official NQF policy.

In all, the SMP agreed that the Landis and Koch scale should no longer be considered an evaluation guide for reliability measurement. SMP's future work will be devoted to identifying a better guidance. The SMP also believes that measure developers need rigorous criteria/guidance in order to improve their submissions. In the future, the SMP urged developers to pay greater attention to misclassification, especially as a metric of reliability, as illustrated by work by Adams, Zaslavsky, and others.

Acceptable Testing Methods for Validity

The SMP highlighted a number of issues regarding validity. Regarding the directionality of process-outcome associations, the SMP does not believe that demonstration of a relationship between a process quality measure and a worse outcome is appropriate. However, NQF guidance does not make this directionality explicit. Therefore, it cannot hold measures to this standard until the guidance is changed through NQF's process. SMP members also highlighted that many measures do not include sufficient hypotheses as to why such correlations between the measure being evaluated and existing measures are important and appropriate. They restated that the description of hypotheses is an essential component of measure submissions and is requested on the testing form. Having hypotheses from developers would build in directionality so that the SMP can tell whether the expected correlation occurs at the right direction. Different validity tests can be used to triangulate an evaluation of the hypothesis.

Some SMP members also pointed out that measure developers should not test the correlation of their measure with aggregate metrics or programs (e.g. Star Ratings) that already contain the measure being discussed, as a correlation is automatic and not meaningful in establishing validity.

Regarding situations where a developer only reports correlations, an ecological fallacy problem can arise. However, these relationships are generally addressed in the evidence section of submissions and so the SMP cannot fully evaluate them. Finally, the SMP agreed that measure developers should say explicitly whether they are assuming a formative or reflective model of measurement in creating composite measures, which affects type of reliability testing and choice of validity measures.

Measure Evaluation

Measure evaluations during the webinar were based on the preliminary analyses performed by assigned members of the SMP. Each SMP member was assigned to one of three subgroups and each subgroup was assigned eight or nine of the 25 measures being discussed this cycle. Subgroup members then performed in-depth reviews and analyses of their assigned measures. Developers received these preliminary analyses prior to the web meeting and were given an opportunity to submit written

responses to concerns expressed. These responses were provided to the SMP prior to the meeting for review to support their discussion and subsequent voting by the subgroups on the measures during the meeting. One measure was withdrawn from consideration by the developer after preliminary evaluation but prior to the web meeting.

During the meeting, the SMP evaluated reliability and validity for eight measures based on their preliminary analyses and additional information submitted for consideration by the developers. For each measure discussed, NQF staff described the measure, noted the preliminary evaluation ratings of the subgroup, and highlighted the criterion (or criteria) for which there was a lack of consensus and/or major areas of concern. David Nerenz and Christie Teigland, SMP co-chairs, facilitated the remainder of the discussion. Two lead discussants from the subgroup first summarized the primary concerns of the subgroup. Other subgroup members then made additional comments. NQF staff then invited measure developers to provide brief responses to the concerns raised by the subgroup members and to summarize their written response, if provided. Next, the co-chairs invited comments or additional questions from other SMP members, and developers were invited to respond to those additional questions and comments. The subgroup members then voted on the measure, producing the final votes of the SMP for the relevant criteria. Quorum was achieved for all subgroup votes. These votes reflect the final overall assessment of reliability and/or validity by the SMP.

The remaining 17 of the 25 measures evaluated by the SMP in the fall 2020 cycle were not discussed during the meeting because subgroup members reached consensus on the ratings, and the measures were not otherwise pulled for discussion. For these measures, the subgroup's preliminary analyses serve as the final overall assessment of reliability and validity for the Standing Committees' consideration.

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

Subgroup 1

During the meeting, the subgroup discussed three measures (#0505, #1891, and #2515) and revoted on reliability for measure #0505. After further discussion, the subgroup accepted the preliminary analysis decisions for measures #1891 and #2515. The subgroup accepted the preliminary analysis decisions for six measures (#0330, #0506, #2888, #3597, #3596, #2158) without further discussion. The final results for the nine measures evaluated by subgroup 1 are presented below.

0505 Hospital 30-day All-Cause Risk-Standardized Readmission Rate (RSRR) Following Acute Myocardial Infarction (AMI) Hospitalization. (Yale Center for Outcomes Research and Evaluation (CORE) / Centers for Medicare & Medicaid Services (CMS))

Measure Steward/Developer Representatives at the Meeting

Doris Peter and Karen Dorsey

Scientific Methods Panel Votes

- Reliability: H-0; M-5; L-4; I-0 (Consensus Not Reached)
- Validity: H-0; M-8; L-1; I-0 (Pass)

In their preliminary analyses, the SMP noted minor concerns on the validity of this measure and passed it with a moderate rating. Though the SMP recognized the method for reliability testing of NQF #0505 was clear and acceptable, there were concerns about the relatively low reliability shown in the results. Specifically, the concern was that the measure cannot reliably detect differences across hospitals with low 25th and 75th percentiles of 0.33 and 0.66, respectively.

During the meeting, the SMP debated about thresholds and standards for reliability statistics. According to the "Landis scale," this measure's agreement score of 0.42 is considered "moderate," but some SMP members pointed out that the values in that scale are arbitrary and were developed for a different purpose. The measure developer and the SMP members agreed that the lack of defined thresholds creates a challenge for assessing reliability of a measure. At the next advisory meeting, the SMP will continue discussions around a scale or defined thresholds to use when evaluating a measure. There were additional concerns related to the approach to social risk adjustment and that when this measure goes to the Standing Committee, this should be considered and discussed. Several social risk factors were tested but not included in the model. The developers noted they intentionally decided not to include these factors and cited the Office of the Assistant Secretary for Planning and Evolution (ASPE) 2020 report to Congress. Another concern was raised around the c-statistic and whether it should be considered low. Again, with the issue of unclear thresholds, it was argued that the value was acceptable and in range with other measures. The SMP revoted after the discussion and did not reach a consensus on reliability during this meeting. The measure will move forward to be reviewed by the All-Cause Admissions and Readmissions Standing Committee in the fall 2020 cycle.

1891 Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization (Yale CORE / CMS)

Measure Steward/Developer Representatives at the Meeting

Doris Peter and Karen Dorsey

Scientific Methods Panel Votes

- Reliability: H-1; M-5; L-3; I-0 (Pass)
- Validity: H-0; M-7; L-2; I-0 (Pass)

In their preliminary analyses, the SMP found this measure to be reliable and valid. However, one SMP member pulled the measure to discuss how each member came to their conclusion for a reliability rating. During the meeting, some SMP members argued results from the split-sample and signal-to-noise tests were low compared to similar measures. The subsequent suggestion was to remove the Landis scale from consideration in future cycles and continue to discuss a more aggregable scale to take its place. This measure passed the preliminary analyses and will be reviewed by the All-Cause Admissions and Readmissions Standing Committee in the fall 2020 cycle.

2515 Hospital 30-day, All-Cause, Unplanned, Risk-Standardized Readmission Rate (RSRR) Following Coronary Artery Bypass Graft (CABG) Surgery (Yale CORE / CMS)

Measure Steward/Developer Representatives at the Meeting

Doris Peter and Karen Dorsey

Scientific Methods Panel Votes

- Reliability: H-1; M-7; L-1; I-0 (Pass)
- Validity: H-1; M-5; L-3; I-0 (Pass)

In their preliminary analyses, the SMP found this measure to be reliable and valid. Similar to NQF #1891, one SMP member pulled the measure to discuss how each member came to their conclusion for a reliability rating. The discussion for NQF #1891 also applied to NQF #2515 where some SMP members argued the split-sample reliability test result was low (<0.5) compared to similar measures. In their written response, the developer reported that their reliability statistic is considered moderate following

the "Landis scale." This measure passed the preliminary analyses and will be reviewed by the All-Cause Admissions and Readmissions Standing Committee in the fall 2020 cycle.

0330 Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Heart Failure (HF) Hospitalization (Yale CORE / CMS)

Scientific Methods Panel Votes

- Reliability: H-0; M-7; L-1; I-0 (Pass)
- Validity: H-2; M-5; L-1; I-0 (Pass)

Subgroup members found the measure to be reliable and valid. The All-Cause Admissions and Readmissions Standing Committee will evaluate this measure in the fall 2020 cycle.

0506 Hospital 30-Day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Pneumonia Hospitalization (Yale CORE / CMS)

Scientific Methods Panel Votes

- Reliability: H-1; M-7; L-1; I-0 (Pass)
- Validity: H-0; M-8; L-1; I-0 (Pass)

Subgroup members found the measure to be reliable and valid. The All-Cause Admissions and Readmissions Standing Committee will evaluate this measure in the fall 2020 cycle.

2888 Risk-Standardized Acute Admission Rates for Patients with Multiple Chronic Conditions (Yale CORE / CMS)

Scientific Methods Panel Votes

- Reliability: H-7; M-1; L-0; I-0 (Pass)
- Validity: H-3; M-3; L-2; I-0 (Pass)

Subgroup members found the measure to be reliable and valid. The All-Cause Admissions and Readmissions Standing Committee will evaluate this measure in the fall 2020 cycle.

3597 Clinician-Group Risk-Standardized Acute Hospital Admission Rate for Patients with Multiple Chronic Conditions under the Merit-based Incentive Payment System (Yale CORE / CMS)

Scientific Methods Panel Votes

- Reliability: H-5; M-2; L-0; I-1 (Pass)
- Validity: H-0; M-7; L-1; I-0 (Pass)

Subgroup members found the measure to be reliable and valid. The All-Cause Admissions and Readmissions Standing Committee will evaluate this measure in the fall 2020 cycle.

3596 Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Acute Ischemic Stroke Hospitalization with Claims-Based Risk Adjustment for Stroke Severity (Yale CORE / CMS)

Scientific Methods Panel Votes

- Reliability: H-3; M-5; L-0; I-0 (Pass)
- Validity: H-1; M-5; L-2; I-0 (Pass)

Subgroup members found the measure to be reliable and valid. The Neurology Standing Committee will evaluate this measure in the fall 2020 cycle.

2158 Medicare Spending Per Beneficiary (MSPB) - Hospital (Acumen, LLC / CMS)

Scientific Methods Panel Votes

- Reliability: H-7; M-0; L-0; I-0 (Pass)
- Validity: H-1; M-6; L-0; I-0 (Pass)

Subgroup members found the measure to be reliable and valid. The Cost and Efficiency Standing Committee will evaluate this measure in the fall 2020 cycle.

Subgroup 2

During the meeting, the subgroup discussed two measures (3599 and 0230) and revoted on validity for measure #3599. After further discussion, the subgroup accepted the preliminary analysis decisions for measure #0230. The subgroup accepted the preliminary analysis decisions for five measures (0229, #0531, #0468, #1550, #1551) without further discussion. One of the eight original measures evaluated by this group was withdrawn from consideration prior to the meeting. The final results for the seven measures evaluated by subgroup 2 are presented below.

3599 Pediatric Asthma Emergency Department Use (University of California, San Francisco)

Measure Steward/Developer Representatives at the Meeting

Naomi Bardach

Scientific Methods Panel Votes

- Reliability: H-2; M-5; L-0; I-1 (Pass)
- Validity: H-0; M-3; L-2; I-1 (Consensus Not Reached)

The SMP members voiced concerns about overfit or underfit for California or Massachusetts analyses, a 30 percent difference in the R^2 results. Many members also felt there were social risk factors missing from consideration.

The SMP also noted that secondary asthma presentation was identified as a potential confounder for the measure. The developer described the issues they found associated with including secondary asthma: it is not uncommon for rhinitis or respiratory infection to result in asthma exacerbation, but there is a challenge in identifying appropriately the trigger or the condition as the primary rationale for the patient presenting in the ER. The developer noted analyses revealed that fever, influenza, and upper respiratory infection are often listed as first diagnoses and asthma is second. The developer also clarified that the measure includes a “second” diagnosis of asthma and not “secondary” diagnoses. The sample size reduction was close to 50 percent ,if a second diagnosis of asthma was excluded with sensitivity analysis showing similar results. Therefore, the developer argued that inclusion of the second diagnosis is important to the measure. One SMP member suggested that the positioning of the diagnosis is irrelevant and may occur beyond the second diagnosis in the coding, which may be a confounder for the measure. However, the developer noted that research has shown that pediatric patients do not tend to have a lot of diagnoses and so the likelihood that asthma would appear lower down in a long list of diagnoses is unlikely. The SMP encouraged developers to cite research like this within their submissions. The Primary Care and Chronic Illness Standing Committee will evaluate this measure in the fall 2020 cycle.

0230 Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization (Yale CORE / CMS)[2]

Scientific Methods Panel Votes

- Reliability: H-0; M-5; L-3; I-0 (Pass)
- Validity: H-0; M-6; L-1; I-1 (Pass)

Subgroup members found the measure to be reliable and valid. The Cardiovascular Standing Committee will evaluate this measure in the fall 2020 cycle.

0229 Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization (Yale CORE / CMS)

Scientific Methods Panel Votes

- Reliability: H-4; M-4; L-0; I-0 (Pass)
- Validity: H-0; M-6; L-1; I-1 (Pass)

Subgroup members found the measure to be reliable and valid. The Cardiovascular Standing Committee will evaluate this measure in the fall 2020 cycle.

0531 Patient Safety and Adverse Events Composite (IMPAQ International / CMS)

Scientific Methods Panel Votes

- Reliability: H-2; M-5; L-0; I-1 (Pass)
- Validity: H-2; M-4; L-1; I-1 (Pass)
- Composite Construction: H-2; M-4; L-1; I-1 (Pass)

Subgroup members found the measure to be reliable and valid. The Patient Safety Standing Committee will evaluate this measure in the fall 2020 cycle.

0468 Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Pneumonia Hospitalization (Yale CORE / CMS)

Scientific Methods Panel Votes

- Reliability: H-4; M-4; L-0; I-0 (Pass)
- Validity: H-1; M-5; L-1; I-1 (Pass)

Subgroup members found the measure to be reliable and valid. The Patient Safety Standing Committee will evaluate this measure in the fall 2020 cycle.

1550 Hospital-level risk-standardized complication rate (RSCR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) (CMS)

Scientific Methods Panel Votes

- Reliability: H-2; M-6; L-0; I-0 (Pass)
- Validity: H-0; M-6; L-1; I-1 (Pass)

[2] Measures NQF 1891, 0230 and 2515 were pulled by SMP members as illustrative examples to see how SMP members apply reliability criteria. During the meeting, the SMP discussed reliability in general without discussing these three measures specifically.

Subgroup members found the measure to be reliable and valid. The Surgery Standing Committee will evaluate this measure in the fall 2020 cycle.

1551 Hospital-level 30-day risk-standardized readmission rate (RSRR) following elective primary total hip arthroplasty (THA) and/or total knee arthroplasty (TKA) (CMS)

Scientific Methods Panel Votes

- Reliability: H-2; M-5; L-1; I-0 (Pass)
- Validity: H-0; M-7; L-0; I-1 (Pass)

Subgroup members found the measure to be reliable and valid. The Surgery Standing Committee will evaluate this measure in the fall 2020 cycle.

Subgroup 3

During the meeting, the subgroup discussed three measures (#3592, #0141, and #0202) and revoted on validity for measures #0141 and #0202. After further discussion, the subgroup accepted the preliminary analysis decision for measure #3592. The subgroup accepted the preliminary analysis decisions for five measures (#3568, #3567, #1623, #3235, #1893) without further discussion. The final results for the eight measures evaluated by subgroup 3 are presented below.

3592 Global Malnutrition Composite Score (Avalere Health LLC)

Measure Steward/Developer Representatives at the Meeting

Angel Valladares

Scientific Methods Panel Votes

- Reliability: H-2; M-4; L-0; I-2 (Pass)
- Validity: H-0; M-6; L-0; I-2 (Pass)
- Composite: H-2; M-3; L-2; I-1 (Pass)

This measure was pulled for discussion by SMP members. One SMP member highlighted that the process-outcome correlations to establish validity on the measure came out in an unexpected direction. Specifically, successful performance on the composite and its component measures were associated with worse outcomes (e.g. longer stay in hospitals, more readmissions) for patients. Therefore, one SMP member raised concerns that the measure is actually measuring how sick patients are instead of quality of care. The SMP member also raised questions that the direction of the correlations between certain component with the outcome is not the always the same direction of the correlation between the composite and the outcome. The developer walked through each step and conditions of how each component of the composite was captured by showing a graphic on the measure workflow.

One member raised that although the SMP is charged with evaluating validity of measures, some aspects of validity may be raised in the evidence section of submissions, which is under the charge of the topic area Standing Committees. At the end of this discussion, the subgroup decided to accept the preliminary analysis decisions for this measure, because the written NQF guidance on composite measures only stated that the developer had to show a relationship between process composite and outcome, not a relationship in the "right" direction. This measure will move to the Standing Committee for evaluation.

0141 Patient Fall Rate (American Nurses Association)

Measure Steward/Developer Representatives at the Meeting

Emily Cramer

Scientific Methods Panel Votes[3]

- Reliability: H-0; M-7; L-0; I-1 (Pass)
- Validity: H-0; M-2; L-6; I-1 (Not Pass)

This measure did not pass on validity during preliminary review and the developer provided additional input so was pulled for discussion and revote at the meeting. Regarding threats to validity, the SMP was concerned about the lack of risk adjustment for case-mix within hospital units. Since the measure would report an aggregate score at the hospital level, the SMP questioned whether stratifying at hospital units level capture the actual case-mix. They suggested adjustment for factors such as age, frailty, and types of medications. The SMP was also concerned about the magnitude of the validity testing correlations and the types of measures used for validity testing. The developer added that they had limited access to patient-level factors (e.g., accurate documentation within electronic health records or an adverse event reporting system) and so, this impacted their ability to test validity and to risk adjust the measure. The reliability of claims data was relatively low, but pulling adverse events data, while admittedly more reliable, would reduce the measure's feasibility and so risk adjustment was not incorporated into the measure. If these issues can be addressed, the developer reported that they would like to risk adjust the measure at some point in the future. The developer reported that this is the reason why they stratify the measure by unit type instead.

The SMP also discussed whether the range of the reliability score testing results should be of concern, specifically the lower bound, as this value was lower than the SMP felt is appropriate. One member mentioned that it was possible the small sample sizes contributed to the lower reliability and that the SMP should consider sample size thresholds in the future.

This measure can be pulled by the Patient Safety Standing Committee for discussion.

0202 Falls with Injury (American Nurses Association)

Measure Steward/Developer Representatives at the Meeting

Emily Cramer

Scientific Methods Panel Votes[4]

- Reliability: H-0; M-7; L-0; I-1 (Pass)
- Validity: H-0; M-3; L-5; I-1 (Not Pass)

This measure is a subset of measure #0141. It did not pass on validity during preliminary review and the developer provided additional input. It was pulled for discussion and revote at the meeting. The SMP raised that this measure's relationship to #0141 should result in a large correlation between the measures. Their similarities also necessitate very similar concerns with risk adjustment and validity testing results. Regarding threats to validity, the SMP did not believe that the measure could adequately detect differences across hospitals and they expressed concerns about discriminant validity at both between-unit within a hospital, and between-hospital. Additional concerns around risk adjustment were focused on social risk factors as with #0141 and so the SMP recommended that the Standing Committee discuss this aspect of the measure specifically. There was also a concern over the threshold of a good c-

[3] Reliability was voted on prior to the meeting and so included all subgroup member votes. One subgroup member was not available for the validity re-vote that took place during the meeting.
[4] Reliability was voted on prior to the meeting and so included all subgroup member votes. One subgroup member was not available for the validity re-vote that took place during the meeting.

statistic, as this measure may be considered low, but a threshold has not been established by NQF. This measure can be pulled by the Patient Safety Standing Committee for discussion.

3568 Person-Centered Primary Care Measure (Virginia Commonwealth University School of Medicine)

Scientific Methods Panel Votes

- Reliability: H-2; M-3; L-1; I-2 (Pass)
- Validity: H-0; M-6; L-0; I-2 (Pass)

Subgroup members found the measure to be reliable and valid. The Primary Care and Chronic Illness Standing Committee will evaluate this measure in the fall 2020 cycle.

3567 Hemodialysis Vascular Access: Practitioner Level Long-term Catheter Rate (University of Michigan Kidney Epidemiology and Cost Center)

Scientific Methods Panel Votes

- Reliability: H-1; M-7; L-0; I-0 (Pass)
- Validity: H-1; M-5; L-1; I-1 (Pass)

Subgroup members found the measure to be reliable and valid. The Renal Standing Committee will evaluate this measure in the fall 2020 cycle.

1623 Bereaved Family Survey (Department of Veterans Affairs / Hospice and Palliative Care)

Scientific Methods Panel Votes

- Reliability: H-1; M-7; L-0; I-0 (Pass)
- Validity: H-0; M-6; L-2; I-0 (Pass)

Subgroup members found the measure to be reliable and valid. The Geriatrics and Palliative Care Standing Committee will evaluate this measure in the fall 2020 cycle.

3235 Hospice and Palliative Care Composite Process Measure—Comprehensive Assessment at Admission (Abt Associates / CMS)

Scientific Methods Panel Votes

- Reliability: H-5; M-3; L-0; I-0 (Pass)
- Validity: H-2; M-5; L-1; I-0 (Pass)
- Composite Construction: H-2; M-6; L-0; I-0 (Pass)

Subgroup members found the measure to be reliable and valid. The Geriatrics and Palliative Care Standing Committee will evaluate this measure in the fall 2020 cycle.

1893 Hospital 30-Day, all-cause, risk-standardized mortality rate (RSMR) following chronic obstructive pulmonary disease (COPD) hospitalization (Yale CORE / CMS)

Scientific Methods Panel Votes

- Reliability: H-0; M-6; L-1; I-0 (Pass)
- Validity: H-2; M-5; L-0; I-0 (Pass)

Subgroup members found the measure to be reliable and valid. The Patient Safety Standing Committee will evaluate this measure in the fall 2020 cycle.

Public Comment

NQF received two public comments through email and chat box during the web meeting.

The first commenter noted a number of considerations if reliability thresholds can be established by NQF. Namely, that methods used to calculate reliability, consideration of case volume for the specific measure, cost, and the arbitrary nature of empirically based thresholds:

- Appropriate interpretation of reliability results requires consideration of the method used to calculate reliability.
- Reliability must also be considered in the broader context of other aspects of a measure's scientific acceptability and overall importance for improving healthcare.
- Further, if NQF decides to pursue reliability thresholds, NQF will need to consider whether they seek to establish a minimum threshold or an averaged or representative threshold for measured entity reliability and provide thresholds for multiple different methods across multiple different measure types and provider types.
- Additionally, any threshold should incorporate the type of statistical distribution associated with the measure.

The second commenter provided a comment regarding the publicly available information on measure NQF #3568 that the signal-to-noise reliability does not appear to assess beyond individual patient response tendency, and there is no publicly available information on case-mix adjustment even though it is categorized as an outcome measure.

Next Steps

NQF staff will inform developers and Standing Committees of the SMP discussion and votes. Measures that passed for both reliability and validity or were consensus not reached will be considered by the relevant standing committees in the fall 2020 evaluation cycle. According to NQF endorsement guidance, measures that did not pass the SMP vote may be pulled for discussion by the relevant Standing Committee. Measures #0141 and #0202 are eligible for the Standing Committee's discussion and revote if they choose. Endorsement may be removed for maintenance measures that did not pass the SMP vote.

The SMP will reconvene via webinar on December 8, 2020 to discuss recommendations for clarifying and updating the current NQF criteria and guidance for reliability and validity testing.