NATIONAL QUALITY FORUM

Meeting Summary

Scientific Methods Panel Measure Evaluation In-Person Meeting

The National Quality Forum (NQF) convened the full Scientific Methods Panel (SMP) for a discussion of the scientific properties (reliability and validity) of measures evaluated during the fall 2019 evaluation cycle on October 28-29, 2019. The 15 measures discussed during this meeting included those for which subgroup 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 NQF staff or subgroup members.

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 lack of consensus and/or major areas of concern. David Cella and David Nerenz, SMP co-chairs, facilitated the remainder of the discussion, wherein a lead discussant from the subgroup that first evaluated the measure noted the primary concerns of the subgroup. Other subgroup members made additional comments. The SMP co-chair then invited measure developers to provide a brief response to the concerns raised by the subgroup members. Next, the co-chair invited comments from other SMP members. The subgroup members who provided an in-depth preliminary analysis of the measure voted on the measure via the Poll Everywhere platform. These votes reflect the final overall assessment of reliability and/or validity by the SMP.

NQF also gave those SMP members who did not provide an in-depth preliminary analysis of the measure an opportunity to vote on the measures offline via the SurveyMonkey platform. This "shadow vote" was conducted as an information-gathering exercise only. It allowed NQF to better understand the scope of the members' review of submission materials, their comfort with providing votes given their level of interaction with the submission materials, and the similarity of their voting responses to those from the subgroup who conducted in-depth evaluations of the measure. The "shadow vote" results are meant for NQF internal use only and do not reflect the official assessment of reliability and/or validity by the SMP.

The remaining seven of the 22 measures evaluated by the SMP in the fall 2019 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 majority recommendations from the subgroup preliminary analyses will serve as the final overall assessment of reliability and validity.

After measure evaluation discussions, the SMP reflected on the new processes incorporated into the fall 2019 cycle and discussed some of the methodological issues that arose over the course of the deliberations.

Welcome, Introductions, and Disclosures of Interest

Karen Johnson, NQF Senior Director, welcomed the members of the Panel. Scientific Methods Panel co-chairs, David Nerenz and David Cella, also provided opening remarks and welcomed returning and new Panel members to the meeting. Elisa Munthali, NQF Senior Vice President,

Quality Measurement, asked Panel members to introduce themselves and provide any disclosures of interest relevant to the measures to be discussed during the meeting. Ms. Johnson then described the process for the measure discussions and reviewed relevant NQF evaluation criteria.

Measure Evaluation

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

Subgroup 1

Subgroup 1 discussed six measures (2456, 1623, 0575, 0059, 0061, and 0425) and accepted the preliminary analysis decisions for one measure (2651) without further discussion. The results for the seven measures evaluated by Subgroup 1 are presented below.

Overarching Issue: Comprehensive Diabetes Care Measures

Subgroup 1 briefly discussed three measures of diabetes care (0575, 0059, and 0061). The measures were found to be reliable and valid in the subgroup's preliminary analyses, but nonetheless, they were pulled for discussion regarding a common issue. The Panel asked the developer to consider the inherent similarities in the measures and explore their potential as a composite. The measure developer (NCQA) noted that there is both an NQF-endorsed composite measure *Optimal Diabetes Care* (NQF 0729), stewarded by Minnesota Community Measurement, as well as NCQA's own composite measure *Comprehensive Diabetes Care* (NQF 0731), which is no longer NQF-endorsed. The Panel also expressed concern that the three measures draw on multiple data sources, but a comparative analysis of the performance by data source was not provided. The Panel then urged the developer to carefully consider the impact of social risk on scoring and performance on the measures. The Panel was not convinced by the developer's argument against the need for risk adjustment and emphasized that many social risk factors may predispose certain populations to have lower performance rates on diabetes-related intermediate outcome measures.

Measures Discussed by the Subgroup

2456 Medication Reconciliation: Number of Unintentional Medication Discrepancies per Patient

Brigham and Women's Hospital

Scientific Methods Panel Votes: Consensus not reached

- Reliability: H-0; M-4; L-2; I-0
- Validity: H-0; M-3; L-2; I-1

In their preliminary analyses, subgroup members did not reach consensus on the validity of the measure. During the Panel's discussion of the measure, members suggested that there may be a need to incorporate some type of additional risk or case-mix adjustment. The Panel noted that it may be more difficult to reconcile medications for patients with more complex regimens (i.e., more medications), leading to a higher likelihood of discrepancies in those patients. While the measure does account for that issue by counting the number of discrepancies per medication per patient, Panel members suggested that the relationship between number of medications and complexity may not be entirely linear, meaning that the developer's approach may not

adequately capture differences in risk across patients. In addition, the developer noted during the discussion that the measure is intended for internal quality improvement purposes, and may not be appropriate for between-hospital comparisons. This caused concern among Panel members, since NQF endorsement implies that measures are suitable for both quality improvement and accountability applications. Ultimately, the subgroup did not reach consensus on the validity of the measure. The measure will have an opportunity to be discussed during the Patient Safety Standing Committee measure deliberations for the fall 2019 cycle.

1623 Bereaved Family Survey

Department of Veterans Affairs / Hospice and Palliative Care

Scientific Methods Panel Votes: Measure does not pass

- Reliability: H-3; M-2; l-0; I-1
- Validity: H-0; M-1; L-4; I-0

In their preliminary analyses, subgroup members did not pass the measure on validity. During their discussion, Panel members expressed concern about the results of the construct validity testing. Specifically, they pointed to the low magnitude of association between the results of the measure and those from four other process measures that reflect provision of high-quality palliative care. Panel members also were confused about whether and how low- versus high-complexity facilities were accounted for in the risk-adjustment approach. Panel members also noted additional concerns regarding the risk-adjustment approach, including the lack of a rationale as to why co-morbidity factors were included and lack of adjustment for sociodemographic factors. Information for this measure and the SMP's deliberations will be provided to the Geriatrics and Palliative Care Standing Committee for the fall 2019 cycle. This measure is eligible for potential revote on validity by the Standing Committee, should one of its members pull it for discussion and/or revote.

0575 Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Control (<8.0%)

National Committee for Quality Assurance

Scientific Methods Panel Votes: Measure passes

- Reliability: H-1; M-4; L-0; I-0
- Validity: H-2; M-2; L-0; I-1

The subgroup members achieved consensus on reliability and validity in their preliminary analyses with a vote on reliability and validity of moderate for both. This measure was discussed in conjunction with measures NQF 0059 and NQF 0061 as summarized above. The Panel elected to retain the vote captured before the meeting after the discussion. The Primary Care and Chronic Illness Standing Committee will evaluate this measure in the fall 2019 cycle. The Panel agreed that a reconsideration of the measure was not warranted, and the votes submitted for the preliminary analysis will stand as the final vote.

0059 Comprehensive Diabetes Care: Hemoglobin A1c (HbA1c) Poor Control (>9.0%)

National Committee for Quality Assurance

Scientific Methods Panel Votes: Measure passes

- Reliability: H-2; M-3; L-0; I-0

- Validity: H-1; M-3; L-0; I-1

The subgroup members achieved consensus on reliability and validity in their preliminary analyses with a vote on reliability and validity of moderate for both of these measures. The Panel noted the similarity between this measure and measure 0575 and retained the vote for Scientific Acceptability of the measure from the preliminary analysis. The Primary Care and Chronic Illness Standing Committee will evaluate this measure in the fall 2019 cycle.

0061 Comprehensive Diabetes Care: Blood Pressure Control (<140/90 mm Hg)

National Committee for Quality Assurance

Scientific Methods Panel Votes:

- Reliability: H-2; M-3; L-0; I-0
- Validity: H-3; M-1; L-0; I-1

In their preliminary analysis, the subgroup members achieved consensus on reliability and validity by a vote of moderate on reliability and a vote of high for validity. This measure was discussed along with measures 0059 and 0575. The Panel retained the vote for Scientific Acceptability of the measure from the preliminary analysis. The Primary Care and Chronic Illness Standing Committee will evaluate this measure in the fall 2019 cycle.

0425 Functional Status Change for Patients with Lumbar Impairments

Focus on Therapeutic Outcomes, Inc (FOTO)

Scientific Methods Panel Votes:

- Reliability: H-3; M-1; L-0; I-1
- Validity: H-4; M-1; L-0; I-0

The subgroup members achieved consensus on reliability and validity in their preliminary analyses with a vote of high on both reliability and validity. This measure was nonetheless pulled for discussion due to concerns that some of the tests performed by the measure developer may have had too small of a sample size and missing data. The measure developer, FOTO, offered explanations for why and how the analyses were conducted. The Panel accepted this explanation, and the measures were moved forward without a revote. The Patient Experience and Function Standing Committee will evaluate this measure in the fall 2019 cycle.

Measure Not Discussed by the Subgroup

2651 CAHPS® Hospice Survey (experience with care)

Centers for Medicare & Medicaid Services

Scientific Methods Panel Votes: Measure passes

- Reliability: H-2; M-4; L-0; I-0
- Validity: H-0; M-6; L-0; I-0

Subgroup members found the eight PRO-PMs included under 2651 to be reliable and valid. The Geriatrics and Palliative Care Standing Committee will evaluate these measures in the fall 2019 cycle.

Subgroup 2

During the meeting, the subgroup discussed four measures (0696, 3537, 0018, and 3534). The subgroup accepted the preliminary analysis decisions for one measure (0071) without further discussion. The final results for the five measures evaluated by subgroup 2 are presented below.

Measures Discussed by the Subgroup

0696 STS CABG Composite Score

The Society of Thoracic Surgeons

Scientific Methods Panel Votes: Measure passes

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

In their preliminary analyses, subgroup members did not reach consensus on validity. During their discussion of the measure, Panel members questioned whether the measure could identify meaningful differences in performance between providers. Some members also questioned the methodology used to demonstrate validity of the measure. Specifically, some did not accept the use of a stability analysis based on star-rating results. Others, however, agreed that identifying an external measure that could be used in validation is challenging, given that the measure itself includes 11 underlying NQF-endorsed measures of mortality and morbidity. Ultimately, subgroup members accepted updated testing based on 2018 data, even though the developers used the same validation methodology. The Surgery Standing Committee will evaluate this measure in the fall 2019 cycle.

3537 Intraoperative Hypotension among Non-Emergent Noncardiac Surgical Cases

Mathematica

Scientific Methods Panel Votes: Measure does not pass

- Reliability: H-2; M-3; L-1; I-0
- Validity: H-0; M-2; L-4; I-0

In their preliminary evaluation, subgroup members did not reach consensus on reliability or validity. Members requested additional detail on the results of reliability testing, which were provided by the developer before the meeting. During the full Methods Panel discussion, members expressed concern about the measure's risk-adjustment approach, noting that some risk variables may be susceptible to gaming (e.g., "up-coding" of ASA risk status to make it appear that patients are at higher risk for interoperative hypotension than they actually are), and that the length-of-surgery risk factor may be impacted by quality of care—i.e., not present before the start of care—and therefore is inappropriate as a risk-adjustment variable. The Panel suggested that using average length of surgery by procedure type may be a more appropriate risk factor. Ultimately, the measure did not pass the Validity criterion. Information for this measure and the SMP's deliberations will be provided to the Patient Safety Standing Committee for the fall 2019 cycle. This measure is eligible for discussion, should one of the Standing Committee members pull it for discussion, but it is NOT eligible for revote on validity by the Standing Committee.

0018 Controlling High Blood Pressure

National Committee for Quality Assurance

Scientific Methods Panel Votes: Measure passes

- Reliability: H-4; M-1; L-0; I-2
- Validity: H-0, M-4, L-2, I-0

In their preliminary evaluation, the subgroup voted to pass the measure for reliability; however, members did not reach consensus on validity. For their evaluation of validity, the subgroup noted concerns with the developer's demonstration of construct validity, which used a comparator measure that was not considered an independent measure of quality. The subgroup also expressed concern with the lack of risk adjustment and lack of testing for the multiple data sources indicated in the specifications.

The developers submitted additional analysis to demonstrate construct validity using a different comparator measure; the SMP agreed that this was a better demonstration of construct validity. The developers also explained that while they did not risk-adjust this measure, they have stratified by health plan type as a proxy for socioeconomic status. Nonetheless, they acknowledged this stratification does not account for clinical differences in patients across plans. They report that this measure is not risk adjusted as it is intended to serve as a population level measure for health plans to determine blood pressure control for their members, regardless of these clinical differences. The SMP ultimately voted to pass this measure on validity. The Prevention and Population Health Standing Committee will evaluate this measure in the fall 2019 cycle.

3534 30 Day All-cause Risk Standardized Mortality Odds Ratio following Transcatheter Aortic Valve Replacement (TAVR)

American College of Cardiology

Scientific Methods Panel Votes: Measure passes

- Reliability: H-0; M-6; L-0; I-0
- Validity: H-0; M-5; L-1; I-0

In their preliminary analyses, subgroup members did not pass the measure on reliability and did not reach consensus on validity. The subgroup's initial concern regarding reliability was lack of testing of many of the critical data elements included in the measure. The subgroup's initial concerns regarding validity focused on the exclusion of more than half of hospitals and patients in the testing sample because of missing data, lack of testing of many of the critical data elements included in the measure, and relatively low kappa values for two data elements. Prior to the evaluation meeting, the measure developers provided additional testing information for the Panel's consideration. This additional information included kappa statistics from inter-rater reliability analysis of several additional data elements for a more recent time period, as well as results of validity testing for several additional data elements for a more recent time period. After discussion of these additional reliability and validity testing results, the subgroup agreed that the measure meets NQF requirements for both reliability and validity. However, the Panel noted that the inclusion of the health status and gait speed variables in the risk-adjustment approach, without imputation for missing values, excludes many hospitals from the measure.

Members therefore recommended that the developers reconsider including these factors. The Cardiovascular Standing Committee will evaluate this measure in the fall 2019 cycle.

Measure Not Discussed by the Subgroup

0071 Persistence of Beta-Blocker Treatment After a Heart Attack
National Committee for Quality Assurance

Scientific Methods Panel Votes: Measure passes

- Reliability: H-2; M-5; L-0; I-0
- Validity H-0; M-5; L-1; I-1

Subgroup members found the measure to be reliable and valid. The Cardiovascular Standing Committee will evaluate this measure in the fall 2019 cycle.

Subgroup 3

Subgroup 3 discussed two measures (3478 and 3492) during the meeting and accepted the preliminary analysis decisions for two measures (0684 and 3538) without further deliberation. The final results for the four measures evaluated by subgroup 3 are presented below.

Measures Discussed by the Subgroup

3478 Surgical Treatment Complications for Localized Prostate Cancer
Alliance of Dedicated Cancer Centers

Scientific Methods Panel Votes: Measure does not pass

- Reliability: H-0; M-2; L-2; I-0
- Validity: H-0; M-1; L-3; I-0

In the preliminary analysis, subgroup members were unable to reach consensus on this measure for both reliability and validity. The subgroup had several concerns with reliability, including the lack of clarity in the specifications on exclusions, attribution, and the transformation of the raw score to the scale reported for the measure score. For the validity criterion, the SMP members expressed concern regarding the lack of risk adjustment that would account for co-occurring conditions and the intent of the measure to use a count of claims as a proxy for surgical outcomes. While the developer acknowledged that complications or co-occurring conditions could have an impact on the measure, they were unable to obtain the data to include this in the measure. The developer clarified the specifications and explained their method for converting the raw score to the values used for reporting the measure. Ultimately, however, the SMP did not believe the developer adequately demonstrated validity of the measure because the numerator relies on counting the number claims before versus after prostate surgery as a proxy for quality of surgical care. The SMP agreed that the number of claims prior to or after surgery can be attributed to many other factors that may not directly relate to the quality of the surgery. It is largely for this reason that the SMP subgroup did not pass this measure on validity; consensus was not reached on reliability. Information for this measure and the SMP's deliberations will be provided to the Cancer Standing Committee for the fall 2019 cycle. This measure is eligible for potential revote on reliability and validity by the Standing Committee, should one its members pull it for discussion and/or revote.

3492 Acute Care Use Due to Opioid Overdose

Yale CORE/Centers for Medicare & Medicaid Services

Scientific Methods Panel Votes: Measure does not pass

- Reliability: H-1; M-2; L-1; I-0
- Validity: H-0; M-1; L-2; I-1

In their preliminary analyses, subgroup members did not reach consensus on the reliability or validity of the measure. Concerns that led to these ratings were largely about the specification of the measure in Medicare Part A and B individuals exclusively, rather than broader assessment of all-payer populations. Validity testing specifically was identified as a concern because there was some reliance on a narrow advisory committee composed only of personnel from the developer's home institution, and there was generally an interest in receiving more details about the empirical validity testing that was conducted. Moreover, the developers did not do risk adjustment to generate their measure even as they said that sociodemographic factors exogenous to the healthcare system influence the rates being measured.

During the in-person meeting, concern persisted about the narrow data scope (Medicare A and B enrollees), the geographic attribution, the face validity composition and description, and the absence of exclusions (i.e., for hospice) composing this measure.

Developers provided score level empirical validity testing results showing that this emergency department overdose measure correlated reasonably with two separate measures of opioid-related death in one case, or hospitalization in a second case. However, subgroup members noted that these comparisons were not limited to the same Medicare population and the same type of events. Some SMP members, although not all, expressed concern regarding the lack of risk adjustment for the measure, particularly geographical adjustment, given the variation in in opioid use in different parts of the U.S.

Information for this measure and the SMP's deliberations will be provided to the relevant Standing Committee for the fall 2019 cycle. This measure is eligible for potential revote on validity by the Standing Committee, should one of its members pull it for discussion and/or revote.

Measures Not Discussed by the Subgroup

0684 Percent of Resident with a Urinary Tract Infection (Long Stay)

Centers for Medicare & Medicaid Services

Scientific Methods Panel Votes: Measure passes

- Reliability: H-0; M-5; L-1; I-0
- Validity: H-1; M-3; L-1; I-1

Subgroup members found the measure to be reliable and valid. The Patient Safety Standing Committee will evaluate this measure in the fall 2019 cycle.

3538 All-Cause Emergency Department Utilization Rate for Medicaid Beneficiaries Who May Benefit from Integrated Physical and Behavioral Health Care

Centers for Medicare & Medicaid Services

Scientific Methods Panel Votes: Measure passes

- Reliability: H-5; M-1; L-0; I-0
- Validity: H-2; M-4; L-0; I-0

Subgroup members found the measure to be reliable and valid. The Behavioral Health and Substance Use Standing Committee will evaluate this measure in the fall 2019 cycle.

Subgroup 4

Subgroup 4 discussed three measures (3528, 3483, 3484) and accepted the preliminary analysis decisions for three measures (2979, 3533e, 3543) without further discussion. The final results for the six measures evaluated by subgroup 4 are presented below.

Measures Discussed by the Subgroup

3528 CDC and VON Harmonized Outcome Measure for Late Onset Sepsis and Meningitis in Very Low Birthweight Neonates

Centers for Disease Control and Prevention

Scientific Methods Panel Votes: Measure does not pass

- Reliability: H-0; M-1; L-6; I-0
- Validity: H-0; M-4; L-2; I-0

In their preliminary analysis, the SMP did not pass this measure on reliability. The principle reason for the reliability failure seemed relate to the fact that the reliability testing was element level limited to just the numerator (infections). Additionally, the testing for reliability seemed more like validity because it was a comparison of manual record reviews to automated record reviews using the developers' so-called "on-line calculator."

The SMP's discussion focused largely on concerns that the specifications for this measure were not clearly described. Specifically, the developer described potentially four different measures in the submission: (1) crude rates, (2) monthly rates, (3) overall survival, and (4) standardized infection ratio (SIR); however, the precise definitions for these calculations were not evident to many of the members who reviewed the submission, including the online calculator described in the reliability testing. Other concerns identified by the reviewers focused on the lack of risk model calibration and fit statistics and the inadequacy of the data element validity testing, which did not suffice to provide confidence in the reliability of the measure given the concerns with the specifications.

Information for this measure and the SMP's deliberations will be provided to the Perinatal Standing Committee for the fall 2019 cycle. This measure is eligible for discussion, should one of the Standing Committee members pull it for discussion, but it is NOT eligible for revote by the Standing Committee.

3483 Adult Immunization Status

National Committee for Quality Assurance

Scientific Methods Panel Votes: Measure passes

- Reliability: H-4; M-1; L-0; I-1
- Validity: H-2; M-3; L-1; I-0
- Composite: H-5; M-0, L-1, I-0

In their preliminary analyses, subgroup members found this measure to be reliable and valid. SMP members pulled this measure for discussion, along with 3483, to discuss the developer's selection of "Integrated Delivery System" as a level of analysis, along with "health plan" even though testing was only submitted for health plan level of analysis. The developer pointed out that their testing was conducted among health plans as well as integrated delivery systems that also served as health plans. SMP members pointed out that not all integrated delivery systems are health plans and are a distinct entity that should be tested if the developers seek to endorse the measure at this level. Acknowledging this, the developers agreed to remove the selection of integrated delivery system for level of analysis.

SMP members raised a second issue: the developer's reliability testing results indicate a nearly perfect reliability score (0.999) for some health plans, using the beta binomial approach (i.e., Adams' method). One SMP member pointed out that the application of this equation at the health plan level, rather than the patient level, contributed to some inflation of the score; however, given the large sample size, this overestimation would likely not have a significant impact on the reliability score, which would still be relatively high. As such, the SMP agreed that it did not need to reconsider the measures and the vote from their preliminary analyses should stand. The Prevention and Population Health Standing Committee will evaluate this measure in the fall 2019 cycle.

3484 Prenatal Immunization Status

National Committee for Quality Assurance

Scientific Methods Panel Votes: Measure passes

- Reliability: H-4; M-1; L-0; I-1
- Validity: H-2; M-3; L-1; I-0
- Composite: H-5; M-0, L-1, I-0

In their preliminary analyses, subgroup members found this measure to be reliable and valid. However, SMP members pulled it for discussion, along with 3483, for the same reasons discussed in the summary of 3483. See the summary for 3483 for details. The SMP did not revote on this measure. The Prevention and Population Health Standing Committee will evaluate this measure in the fall 2019 cycle.

Measures Not Discussed by the Subgroup

2979 Standardized Transfusion Ratio for Dialysis Facilities

Centers for Medicare & Medicaid Services

Scientific Methods Panel Votes: Measure passes

- Reliability: H-2; M-3; L-1; I-0
- Validity: H-4; M-0; L-0; I-2

Subgroup members found the measure to be reliable and valid. The Renal Committee will evaluate this measure in the fall 2019 cycle.

3533e Hospital Harm – Severe Hyperglycemia

IMPAQ International LLC

Scientific Methods Panel Votes: Measure passes

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

Subgroup members found the measure to be reliable and valid. The Safety Committee will evaluate this measure in the fall 2019 cycle.

3543 Patient-Centered Contraceptive Counseling (PCCC) measure

UCSF

Scientific Methods Panel Votes: Measure passes

- Reliability: H-5; M-1; L-0; I-0
- Validity: H-5; M-1; L-0; I-0

Subgroup members found the measure to be reliable and valid. The Perinatal and Women's Health Standing Committee will evaluate this measure in the fall 2019 cycle.

Process Review: Fall 2019 Cycle

The SMP co-chairs led the Panel in discussing several of the changes to the SMP process during the fall 2019 cycle to date. Panel members supported the switch to an in-person meeting format, allowing all Panel members to discuss the measures, and allowing measure developers to submit additional information prior to the in-person meeting. Members also expressed appreciation for the discussion scripts provided by NQF staff, although they suggested some improvements. Staff also shared overall results of the shadow votes with the SMP. Of the 13 votes conducted during the meeting, 11 of the shadow vote results matched the results from the subgroup votes. SMP members were not surprised with the extent of agreement between the votes. NQF will continue to review the voting process for SMP members and implement improvements as needed in future evaluation cycles.

Discussion of Methodological Issues Identified During Measure Evaluation

The SMP also discussed potential changes to NQF's requirements for testing for reliability and validity. Across the board, members recommended that NQF discontinue its current practice of waiving demonstration of reliability when adequate data element validation has been demonstrated. Members also agreed that NQF should require score-level reliability testing for all measures submitted for potential endorsement. Members did not agree on whether score-level validity testing should be required for all measures. However, some favored this requirement if an exception for not conducting such testing could be granted, assuming an adequate rationale is provided. Similarly, members did not agree on the need for data element

reliability and validity testing for all measures, although again, some favored these requirements if an exception could be granted when adequate justification is provided.

Next Steps

NQF Consensus Development Process (CDP) project teams will inform developers and standing committees of the SMP discussion and votes. Measures rated as high or moderate for both reliability and validity will be evaluated by the relevant standing committees in the fall 2019 evaluation cycle. Measures that did not pass SMP vote may be pulled for discussion by the relevant standing committee. NQF staff, in consultation with SMP co-chairs, will determine if measures that did not pass the SMP vote are eligible for revote by the standing committee in the fall 2019 evaluation cycle.