NATIONAL QUALITY FORUM

Memo

May 9, 2019

To: All-Cause Admissions and Readmissions 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 All-Cause Admissions and Readmissions Standing Committee will meet via web meeting on May 16, 2019 from 12:00 pm to 2:00 pm ET. The purpose of this call is to:

- Review and come to consensus on a recommendation for 3366;
- 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' expression of support of the measures under consideration;
- Determine whether reconsideration of any measures or other courses of action are warranted; and
- Discuss a request for reconsideration on 3443 and 3445

Standing Committee Actions

1. Review this briefing memo and the draft report.
2. Review and consider the full text of all comments received and the proposed responses to the post-evaluation comments (see comment table and additional documents included with the call materials).
3. Review the NQF members' expressions of support of the submitted measures.
4. Review the request for reconsideration submitted on 3443 and 3445.
5. Be prepared to provide feedback and input on proposed post-evaluation comment responses.
6. Be prepared to vote on a recommendation for endorsement for 3366.

Conference Call Information

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

Dial-in #: 800-768-2983
Passcode: 2861387
Web link: https://core.callinfo.com/callme/?ap=8007682983&ac=2861387&role=p&mode=ad

Background

Quality improvement has a critical goal of reducing avoidable hospital admissions and readmissions. Avoidable admissions and readmissions take patients away from their daily lives

and contribute to unnecessary healthcare spending. However, concerns about the unintended consequences of using measures of admissions and readmissions in accountability programs have prompted important study and discussion about how to meet quality goals while protecting access to necessary and appropriate care. NQF currently has 50 endorsed all-cause and condition-specific admissions and readmissions measures addressing numerous settings. Several federal quality improvement programs have adopted these measures to reduce unnecessary admissions and readmissions by fostering improved care coordination across the healthcare system.

On February 7, 2019, the 25-member All-Cause Admissions and Readmissions Standing Committee met to evaluate seven newly submitted measures against NQF's standard evaluation criteria. The Committee recommended three measures for endorsement, did not reach consensus on one measure, and did not recommend three measures. The Standing Committee recommended the following three measures for endorsement:

- 3449 Hospitalization for Ambulatory Care Sensitive Conditions for Dual Eligible Beneficiaries
- 3457 Minimizing Institutional Length of Stay
- 3470 Hospital Visits after Orthopedic Ambulatory Surgical Center Procedures

The Committee did not reach consensus on the following measure:

- 3366 Hospital Visits after Urology Ambulatory Surgical Center Procedures

The Committee did not recommend the following measures:

- 3443 All-Cause Emergency Department Utilization Rate for Medicaid Beneficiaries with Complex Care Needs and High Costs (BCNs)
- 3445 All-Cause Inpatient Admission Rate for Medicaid Beneficiaries with Complex Care Needs and High Costs (BCNs)
- 3456 Admission to an Institution from the Community

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 5, 2018 to January 25, 2019 for the measures under review. No comments were submitted prior to the Standing Committee's evaluation meeting. Therefore, the Committee did not consider any comments 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 16, 2019 for 30 calendar days. During this commenting period, NQF received nine comments from four member organizations:

Member Council	# of Member Organizations Who Commented
Health Professional	2
Provider Organization	1
QMRI	1

We have included all comments that we received in the comment table (excel spreadsheet) posted to the Committee SharePoint site. This comment table contains the commenter's name, comment, associated measure, topic (if applicable), and draft responses (including measure steward/developer responses) for the Committee's consideration. Please review this table before the meeting and consider the individual comments received and the proposed responses to each.

In order to facilitate discussion, the majority of the comments have been categorized into major topic areas or themes. Although all comments are subject to discussion, the intent is not to discuss each individual comment on the May 16, 2019 post-comment call. Instead, we will spend the majority of the time considering the two themes discussed below, and the set of comments as a whole. Please note that the organization of the comments into major topic areas is not an attempt to limit Committee discussion. Additionally, please note that measure stewards/developers were asked to respond where appropriate. Where possible, NQF staff has proposed draft responses for the Committee to consider.

Comments and their Disposition

Themed Comments

Two major themes were identified in the post-evaluation comments, as follows:

1. Adjustment for social risk factors
2. Adequate variation in performance for accountability applications

Theme 1 – Adjustment for Social Risk Factors

Commenters raised concerns about the adequacy of the testing of the impact of social risk factors. Specifically, commenters noted their concern that social risk factors were tested after adjustments were made for clinical risk factors. Bivariate testing of social risk factor testing could provide additional information about how each factor performs. Commenters also suggested that developers continue to test new social risk variables, particularly ones that directly reflect the community in which a patient resides. However, commenters did note their support and continued encouragement for the Standing Committee's discussion on adjustment of social risk factors. Commenters highlighted their agreement with the Committee's discussions around how best to approach adjustment (adjustment versus stratification) as it applies to

different measures intended for different purposes, concerns that developers may hold social risk factors to a higher standard for inclusion in risk models, and deliberations on how to minimize the unintended consequences of measurement for patients.

Proposed Committee Response

The Committee has reviewed your comment and appreciates your input. The Committee agrees that the relationship between social risk factors and patient outcomes is an important area of emerging research. It is critical that developers examine the conceptual relationship between social risk factors and the empirical relationship together. However, the Committee recognizes the challenge developers face in obtaining precise social risk data, which can lead to a discrepancy between the conceptual basis for including social risk factors and the empirical analyses demonstrating their impact. The Committee recognizes that developers may decide about whether to include social risk factors based on whether the factors were related to a provider's quality versus a person's intrinsic risk. However, the Committee also recognizes the need to maximize the predictive value of a risk-adjustment model and ensure that providers serving vulnerable populations are not penalized unfairly.

While the Committee generally accepted the findings of the analyses conducted by the developer, the Committee agrees that more work is needed to identify more robust data elements and methods to isolate and account for unmeasured clinical and social risk for patients. The Committee encourages the developer to continue testing the risk-adjustment model with additional social risk factors to understand their independent contribution to explaining variation in patient outcomes.

Action Item

Does the Committee agree with the proposed response?

Theme 2 – Adequate Variation in Performance for Accountability Applications

Commenters noted the relatively limited amount of variation across applicable ambulatory surgical centers (ASCs) found during testing of 3366 and 3470 and raised concerns about whether these measures provide useful information for accountability and informing patients of the quality of care provided. Commenters noted that endorsing a measure that only identifies a small number of outliers may not provide meaningful information for end users and is inconsistent with the usability and use criterion.

Proposed Committee Response

The Committee has reviewed your comment and appreciates your input. The Committee agrees that these measures demonstrate relatively limited variation across ambulatory surgery centers. However, the Committee believes that the measures provide important information on outliers despite a narrow distribution, and the odds ratios provided may indicate overall less than optimal performance on this measure.

Action Item

Does the Committee agree with the proposed response?

Measure-Specific Comments

3366 Hospital Visits after Urology Ambulatory Surgical Center Procedures

Commenters raised concerns about the validity and usability of this measure. Under the validity subcriterion commenters questioned the lack of adjustment for social risk factors. Specifically, commenters questioned the developer's decision to test the impact of social risk factors after the clinical factors had been added to the model. Concerns about the usability of this measure related to the narrow range of performance across facilities. Commenters questioned if this measure gave useful information for accountability purposes. One commenter questioned why this Committee agreed that measure 3470 was valid but could not reach consensus on this measure.

Measure Steward/Developer Response

We appreciate the opportunity to respond to the concern that the measure is not adequately tested or adjusted for social risk factors (SRFs). We want to clarify the full range of work that we conducted. Guided by NQF standards, findings of the IMPACT Act-funded work of the National Academy of Science, Engineering and Medicine (NASEM), and by the literature, with input from stakeholders and CMS, we developed with stakeholders and presented a conceptual model of how SRFs may influence the outcome. We discuss how socioeconomic factors such as income and health literacy, as well as quality, can influence the measured outcome of hospital visits within 7 days (Sec. 2b3.3a). We then selected three variables based on the literature and NASEM recommendations (dual eligibility [DE], African-American race, and AHRQ SES Index), and tested each variable for its bivariate association with the outcome and its marginal effect on the risk model (after adjusting for other variables). Further, to assess how adjusting or not may affect the measure score, we evaluated the relationship of the measure score at a facility to the proportion of patients with each risk factor at the facility (Sec. 2b3.4b).

In brief, the results (Sec. 2b3.4b) showed that DE patients and those with lower SES status as indicated by the AHRQ SES Index have higher unadjusted rates and are significantly correlated in bivariate analyses with the outcome. In contrast, 7-day hospital visit rates were not higher for African Americans, and the bivariate relationship was not statistically significant. Given the relationship of DE and AHRQ SES status to the outcome, we further examined whether adding the variables (or not adding them) to the risk model was likely to affect the measure score. Adding them to the model did not substantially improve the model c-statistic, suggesting other variables already carried much of the risk. Finally, we examined whether the facility measure scores were higher (worse) for facilities with higher proportions of patients with social risk factors. We found that they were skewed slightly higher, although the difference was not statistically significant. There was variation in performance among those facilities with the highest proportion of low SRF patients, indicating that good performance on the score is achievable for these facilities. We therefore concluded that although dual-eligible patients have higher risk of an event, inclusion of this risk-factor has little to no influence on the measure score.

We appreciate the opportunity to further clarify why, given the conceptual model and the results of this testing, the measure is not adjusted for SES. CMS's decision to not adjust for SES was informed by several factors—the testing results, the conceptual pathways identified by stakeholders and in the literature for how these factors may influence the outcome, and stakeholder input during measure public comment.

CMS and stakeholders considered the tradeoffs inherent in adjusting or not adjusting. There are potential downsides to adjusting for SRFs. If outcomes are systematically worse (higher) for patients with social risk factors and if this is in part due to quality differences, then adjusting could mask quality differences associated with the risk factor. If patients with the risk factor systematically receive poorer quality care (or if known interventions to address social risk factors, such as literacy, are underutilized), and their hospital visit rates are higher as a result, adjusting for the SRF will hinder the measure's ability to drive improvements in care by making such quality differences less visible. On the other hand, there are potential unintended consequences of not adjusting. If certain risk factors strongly influence the outcome in ways unrelated to quality, not adjusting for them could reflect case mix rather than quality. Moreover, if providers anticipate worse outcomes for patients with social risk factors, the measure could create an incentive to reduce access to care for vulnerable patients. Finally, if the measure is used in programs that reduce payment based on the measure scores, not adjusting might reduce resources among the very providers already facing the largest resource constraints; however, this latter concern is not applicable in the ASC Quality Reporting program for which the measure is developed.

CMS weighed these considerations and the test results for this particular measure for this particular program and determined that the downsides to adjustment outweighed the upsides of adjusting given the risk factors did not improve model performance at the margin, the measure score varies among providers with the most patients with SRFs, and because through the measure development process, the TEP and other stakeholders supported not adjusting so as not to mask disparities.

We appreciate the opportunity to clarify why we tested the marginal impact of social risk factors after adjusting for other risk factors such as clinical comorbidity and procedural complexity. As discussed above there are tradeoffs inherent in adjusting for SRFs. Adjusting potentially masks disparities in care, and potentially reduces incentives to address the needs of patients with social risk factors during the provision of care. On the other hand, not adjusting for SRFs that are related to the outcome and cannot practically be mitigated through better care has downsides, including dis-incentivizing care for patients with SRFs. Clinical risk factors don't impose these same tradeoffs. Hence, we tested the marginal effect of SRFs after adjusting for clinical risk factors to inform consideration of these tradeoffs by CMS, experts, and stakeholders. The NQF Scientific Methods Panel members who reviewed the measure noted that the approach used was "thorough and appropriate" and that the "discussion of socio-demographic variables was extensive."

Our current analysis is consistent with the recommendations from ASPE's 2016 report (their latest report is due to be released in October 2019) and with the National Academy of Sciences 2017 report on this same topic, which identified dual eligibility as the most robust variable for SRF adjustment of claims-based measures (see Table 2.1, page 40 of ASPE's 2016 report). The field is evolving, however, and for measures that are in earlier stages of development focused on populations of patients with chronic conditions we are looking at a broader range of variables and assessing their conceptual and statistical relationship to the outcome. For example, CMS recently posted for public comment a Merit-Based Incentive Payment System measure under development of unplanned admissions for patients with multiple chronic conditions. For that measure, in addition to examining DE, we are exploring risk adjusting for several area factors, including rurality and specialist density, identified as potentially influencing outcome rates largely independent of quality. The rationale and variables are currently undergoing public comment, so are available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/PC-Currently-Accepting-Comments.html. This urology measure, however, is focused on a narrower group of patients with literature supporting the types of variables we looked at; area level variables were not raised by stakeholders during development.

In terms of variation, overall, the goal of measurement is to broadly improve quality and narrow the variation in care. What constitutes substantial variation is subjective and contextual; this measure covers outpatient procedures after which patients are not expected to seek hospital care, yet our data show that patients are returning to the hospital for an emergency department (ED) visit, observation stay, or admission after an ASC urology procedure 5.8% of the time. During our Technical Expert Panel, participants indicated that this rate was too high given that the expectation for ASC-based procedures is that patients selected for the procedures will not need follow-up acute care, and that ultimately the goal should be near zero.

There is also meaningful variation in the measure score. As presented in the NQF application, the range of performance on the measure (the risk-standardized hospital visit rate), estimated using Medicare FFS data (FYs 2014-2015) ranged from 3.7% to 10.1% (median of 5.8%). The median odds ratio was 1.27, which represents the median increase in odds of a hospital visit if a procedure on a single patient was performed at a higher risk ambulatory surgery center (ASC) compared to a lower risk ASC. (The median odds ratio is interpreted as a traditional odds ratio would be.)

We appreciate the commenter's point, however, that the measure identified few outliers using conservative 95% confidence intervals. This is not unexpected. The measure's low outcome rate (combined with lower volumes) will reduce the precision of estimates leading to wider confidence intervals. This, however, does not diminish the importance of the measure; we observed many avoidable complications as part of the outcome and substantial variance in both observed and risk-adjusted rates among ASCs. Providing the risk adjusted rates and identifying those facilities that are outliers with a

very high degree of confidence using the 95% CI can be informative to consumers and ASCs.

In summary, the score variation and the relatively high average RSHVR given that returns to the hospital should be relatively unexpected after ASC procedures together show a clear quality gap. The results suggest that there is substantial opportunity to reduce the overall rate and the variation in rates across ASCs, and that this improvement goal is achievable.

Proposed Committee Response

Thank you for your feedback on measure 3366. The Committee will take these comments into account during the post-comment conference call.

Action Item

The Committee did not reach consensus on the measure's validity. The Committee must discuss and vote on the validity subcriterion. If the measure passes validity, the Committee must vote on an overall recommendation for endorsement.

3470 Hospital Visits after Orthopedic Ambulatory Surgical Center Procedures

Comments on measure 3470 were similar to measure 3360. Commenters raised concerns about the validity and usability of this measure. Under the validity subcriterion, commenters questioned the lack of adjustment for social risk factors. Specifically, commenters questioned the developer's decision to test the impact of social risk factors after the clinical factors had been added to the model. Concerns about the usability of this measure related to the narrow range of performance across facilities.

Measure Steward/Developer Response

We appreciate the opportunity to respond to the concern that the measure is not adequately tested or adjusted for social risk factors (SRFs). We want to clarify the full range of work that we conducted. Guided by NQF standards, findings of the IMPACT Act-funded work of the National Academy of Science, Engineering and Medicine (NASEM), and by the literature, with input from stakeholders and CMS, we developed with stakeholders and presented a conceptual model of how SRFs may influence the outcome. We discuss how socioeconomic factors such as income and health literacy, as well as quality, can influence the measured outcome of hospital visits within 7 days (Sec. 2b3.3a). We then selected three variables based on the literature and NASEM recommendations (dual eligibility [DE], African-American race, and AHRQ SES Index), and tested each variable for its bivariate association with the outcome and its marginal effect on the risk model (after adjusting for other variables). Further, to assess how adjusting or not may affect the measure score, we evaluated the relationship of the measure score at a facility to the proportion of patients with each risk factor at the facility (Sec. 2b3.4b).

In brief, the results (Sec. 2b3.4b) showed that DE patients and those with lower SES status as indicated by the AHRQ SES Index have higher unadjusted rates; presence of the DE variable (but not lower SES as indicated by the AHRQ SES Index) was also significantly correlated in bivariate analyses with the outcome. In contrast, 7-day hospital visit rates

were not higher for African Americans, and the bivariate relationship was not statistically significant. Given the relationship of DE status to the outcome, we further examined whether adding the SRF variables (or not adding them) to the risk model was likely to affect the measure score. Adding them to the model did not substantially improve the model c-statistic, suggesting other variables already carried much of the risk. Finally, we examined whether the facility measure scores were higher (worse) for facilities with higher proportions of patients with social risk factors. We found that they were skewed slightly higher, although the difference was not statistically significant. There was variation in performance among those facilities with the highest proportion of low SRF patients, indicating that good performance on the score is achievable for these facilities. We therefore concluded that although dual-eligible patients have higher risk of an event, inclusion of this risk-factor has little to no influence on the measure score.

We appreciate the opportunity to further clarify why, given the conceptual model and the results of this testing, the measure is not adjusted for SES. CMS's decision to not adjust for SES was informed by several factors—the testing results, the conceptual pathways identified by stakeholders and in the literature for how these factors may influence the outcome, and stakeholder input during measure public comment.

CMS and stakeholders considered the tradeoffs inherent in adjusting or not adjusting. There are potential downsides to adjusting for SRFs. If outcomes are systematically worse (higher) for patients with social risk factors and if this is in part due to quality differences, then adjusting could mask quality differences associated with the risk factor. If patients with the risk factor systematically receive poorer quality care (or if known interventions to address social risk factors, such as literacy, are underutilized), and their hospital visit rates are higher as a result, adjusting for the SRF will hinder the measure's ability to drive improvements in care by making such quality differences less visible. On the other hand, there are potential unintended consequences of not adjusting. If certain risk factors strongly influence the outcome in ways unrelated to quality, not adjusting for them could reflect case mix rather than quality. Moreover, if providers anticipate worse outcomes for patients with social risk factors, the measure could create an incentive to reduce access to care for vulnerable patients. Finally, if the measure is used in programs that reduce payment based on the measure scores, not adjusting might reduce resources among the very providers already facing the largest resource constraints; however, this latter concern is not applicable in the ASC Quality Reporting program for which the measure is developed.

CMS weighed these considerations and the test results for this particular measure for this particular program and determined that the downsides to adjustment outweighed the upsides of adjusting given the risk factors did not improve model performance at the margin, the measure score varies among providers with the most patients with SRFs, and because through the measure development process, the TEP and other stakeholders supported not adjusting so as not to mask disparities.

We appreciate the opportunity to clarify why we tested the marginal impact of social risk factors after adjusting for other risk factors such as clinical comorbidity and procedural complexity. As discussed above there are tradeoffs inherent in adjusting for SRFs. Adjusting potentially masks disparities in care, and potentially reduces incentives to address the needs of patients with social risk factors during the provision of care. On the other hand, not adjusting for SRFs that are related to the outcome and cannot practically be mitigated through better care has downsides, including dis-incentivizing care for patients with SRFs. Clinical risk factors don't impose these same tradeoffs. Hence, we tested the marginal effect of SRFs after adjusting for clinical risk factors to inform consideration of these tradeoffs by CMS, experts and stakeholders. The NQF Scientific Methods Panel members who reviewed the measure noted that the approach used was "thorough and appropriate" and that the "discussion of socio-demographic variables was extensive. Our current analysis is consistent with the recommendations from ASPE's 2016 report (their latest report is due to be released in October 2019) and with the National Academy of Sciences 2017 report on this same topic, which identified dual eligibility as the most robust variable for SRF adjustment of claims-based measures (see Table 2.1, page 40 of ASPE's 2016 report). The field is evolving, however, and for measures that are in earlier stages of development focused on populations of patients with chronic conditions we are looking at a broader range of variables and assessing their conceptual and statistical relationship to the outcome. For example, CMS recently posted for public comment a Merit-Based Incentive Payment System measure under development of unplanned admissions for patients with multiple chronic conditions. For that measure, in addition to examining DE, we are exploring risk adjusting for several area factors, including rurality and specialist density, identified as potentially influencing outcome rates largely independent of quality. The rationale and variables are currently undergoing public comment, so are available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/PC-Currently-Accepting-Comments.html. This orthopedic measure, however, is focused on a narrower group of patients with literature supporting the types of variables we looked at; area level variables were not raised by stakeholders during development.

While the outcome rate and the range in measure scores are relatively modest for this measure, orthopedic procedures are common procedures, and the variation shows room for improvement. As presented in the NQF application, performance on the measure (the risk-standardized measure scores), estimated using Medicare FFS data (FYs 2014-2015) ranged from 1.8% to 3.8% (median of 2.5%). The median odds ratio of 1.22 suggests there are meaningful increases in the risk of a hospital visit across facilities. Specifically, if a procedure on the same patient was performed at a higher risk ASC compared to a lower risk ASC the patient on average has a 22% increase in the odds of a hospital visit if the same procedure was performed at higher risk ASC compared to a lower risk ASC.

The median RSHVR of 2.5% is relatively low; however, during measure development our Technical Expert Panel indicated that this rate was too high given that the expectation

for ASC-based procedures is that patients selected for the procedures will not need follow-up acute care, and that ultimately the goal should be near zero.

We appreciate the commenter's point, however, that the measure identified few outliers using conservative 95% confidence intervals. This is not unexpected. The measure's low outcome rate (combined with lower volumes) will reduce the precision of estimates leading to wider confidence intervals. This, however, does not diminish the importance of the measure; we observed many avoidable complications as part of the outcome and substantial variance in both observed and risk-adjusted rates among ASCs. Providing the risk adjusted rates and identifying those facilities that are outliers with a very high degree of confidence using the 95% CI can be informative to consumers and ASCs.

Overall, our results suggest that there is substantial need to both reduce the expected rate and the variation in rates across ASCs, and that this improvement goal is achievable.

Proposed Committee Response

The Committee has reviewed your comment and appreciates your input. The Committee agrees that the relationship between social risk factors and patient outcomes is an important area of emerging research. It is critical that developers examine the conceptual relationship between social risk factors and the empirical relationship together. However, the Committee recognizes the challenge developers face in obtaining precise social risk data, which can lead to a discrepancy between the conceptual basis for including social risk factors and the empirical analyses demonstrating their impact. However, the Committee also recognizes the need to maximize the predictive value of a risk-adjustment model and ensure that providers serving vulnerable populations are not penalized unfairly.

While the Committee generally accepted the findings of the analyses conducted by the developer, the Committee agrees that more work is needed to identify more robust data elements and methods to isolate and account for unmeasured clinical and social risk for patients. It is critical that developers examine the conceptual relationship between social risk factors and the empirical relationship together in an effort to better understand unmeasured patient risk.

The Committee agrees that this measure demonstrates relatively limited variation across ambulatory surgery centers. However, the Committee believes that this measure provides important information on outliers despite a narrow distribution and potentially overall less than optimal performance. Specifically, the Committee notes that the measure developer reported a measure performance range of 1.6 percent to 4.4 percent and a median measure performance of 2.5 percent. Moreover, developers noted a median odds ratio of 1.22 that would suggest that the odds of an unplanned hospital visit are 22 percent higher at a higher-risk ASC versus a lower-risk ASC.

Action Item

Does the Committee agree with the proposed response?

3456 Admission to an Institution from the Community

The measure developer provided several clarifications about the measure and the Standing Committee's deliberations on it. First, the developer noted that the intention of the measure is to reduce unnecessary admissions to nursing homes and other facilities by delivering appropriate long-term services and supports in the community. The developer commented that this concept is important to patients and families and that MLTSS plans can reduce unnecessary admissions by increasing the use and quality of home and community-based services through person-centered assessment, care planning, and care coordination. The developer agreed that a rate of zero on this measure is not desirable or possible but that the measure's intent is to gauge the strength and performance of health plans' ability to provide timely access to high-quality HCBS, not discourage the use of all institutional care.

Additionally, the developer clarified that the measure is designed to compare performance of MLTSS plans within states, not across them and that this may address some of the Committee's concerns about variation among states in data availability and benefit design. The developer noted the measure is specified at the health plan level of analysis and would allow each state to compare the performance of the MLTSS plans with which they are contracting. In addition, the measure will give beneficiaries the chance to compare plan performance when choosing plans in which to enroll.

Thirdly, the developer provided clarifications on the measure's risk-adjustment strategy. Based on the recommendations of its risk-adjustment workgroup and other experts, the developer adopted an age-stratification approach to risk adjustment. They believe this is the best option for this measure in that it provides an easily understandable method for reporting plan performance across relevant age groups.

Finally, the developer provided a response to the Committee's concerns about lowering quality and access. The developer noted that in in most states Medicaid beneficiaries enrolled in managed care plans, including MLTSS plans, are required to enroll in such plans to receive services. Mandatory enrollment does not eliminate the potential for plans to avoid high-risk enrollees (that is, to cherry-pick), but it greatly reduces their ability to engage in such behavior. Additionally, the developer notes that this measure could help identify areas were HCBS services are in short supply, and MLTSS plans can use several proven strategies to improve access to HCBS, thereby improving their performance on this measure. Moreover, the developer notes that lowering rates of institutionalization should not be assumed to lower quality of outcomes and notes that the evidence does not support the assumption that institutionalization has uniformly better effects than HCBS. This measure would allow for within-state plan comparisons that could help states identify best practices in balancing access to HCBS with access to institutions.

Proposed Committee Response

Thank you for your feedback on measure 3456. The Committee will take these developer comments related to the measure intent, design, the risk-adjustment approach, and concerns related to access into account during the post-comment conference call.

Action Item

Does the Standing Committee wish to reconsider its decision on the validity of 3456?

Request for Reconsideration

The measure steward and developer team for measures 3443 and 3445 have requested that the Committee reconsider their decision not to recommend these measures. Appendix B contains the full text of the requests.

3443 All-Cause Emergency Department Utilization Rate for Medicaid Beneficiaries with Complex Care Needs and High Costs (BCNs) and 3445 All-Cause Inpatient Admission Rate for Medicaid Beneficiaries with Complex Care Needs and High Costs (BCNs)

CMS and Mathematica have requested a consideration of the Committee's decision not to recommend 3443 and 3445. The measure steward and developer have responded to the Committee's concerns about the measures' validity. Specifically, they have responded to the Committee's concern about differences in Medicaid populations across states, whether the measure was tested with a representative data sample, and the data quality.

The measure steward and developer provided several clarifications regarding the differences in Medicaid populations across states. They recognized that state Medicaid programs vary substantially both in the covered populations and the quality of data reported to CMS. However, these variations are due to the design of Medicaid of federal-state partnership, and the developer raised concerns that the Committee's emphasis on state variation in Medicaid program created an unrealistic standard for validity.

The steward and developer noted that they believe the measure was tested using a robust data sample for assessing measure performance. They noted the states providing data varied in location, geography, size, and delivery system (fee-for-service or managed care) while still providing high-quality data. Additionally, they commented that the differences across Medicaid programs due to eligibility policies, mix of delivery models, payment rates, and other features, make it challenging for any sample to be representative of all 50 states. The steward and developer commented that the goal of measure testing is to select a diverse group of states that have high-quality data and whose populations capture, for the key variables in question, the majority of the variation that also occurs within other states. They also clarified that the measure specifications were designed to maximize the likelihood that states could define the denominator population consistently.

Finally, the steward and developer responded to the Committee's concerns about data quality. The developer worked to evaluate the quality of relevant data in all the states and selected those states whose data met our quality standards. Specifically, the states chosen for testing had indicators that aligned with national inpatient and emergency utilization benchmarks and did not have data anomalies that would raise analytic problems (such as high levels of missing data). Further, NQF has endorsed measures in the past that were tested with Medicaid data from the same data source. The measure steward and developer noted that they do not believe that state variation in data quality should be a key factor in determining suitability of Medicaid measures for endorsement as long as the data used for development is of sufficient quality. Additionally, they note that even perfect data from all states will not change the fact that state Medicaid programs have differences in design and operational features.

Proposed Committee Response

Thank you for your feedback on measure 3443 and 3445. The Committee will take these comments into account during the post-comment conference call and determine if it wishes to reconsider its recommendation for endorsement.

Action Item

The Committee must determine whether or not to accept the request for reconsideration. If accepted, the Committee must discuss and vote on the validity subcriterion. If the measure passes validity, the Committee must vote on an overall recommendation for 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. Two NQF members provided their expressions of nonsupport for two measures: See Appendix A.

Appendix A: NQF Member Expression of Support Results

Two NQF members provided their expressions of nonsupport on two of the measures. None of the seven measures under consideration received support from NQF members. Results for each measure are provided below.

3366 Hospital Visits after Urology Ambulatory Surgical Center Procedures (CMS/Yale)

Member Council	Support	Do Not Support	Total
Health Professional	0	1	1
Provider Organization	0	1	1
All Councils	**0**	**2**	**2**

3470 Hospital Visits after Orthopedic Ambulatory Surgical Center Procedures (CMS/Yale)

Member Council	Support	Do Not Support	Total
Provider Organization	0	1	1
All Councils	**0**	**1**	**1**

Appendix B: Request for Reconsideration

TO: National Quality Forum and Admissions/Readmissions Standing Committee
FROM: Henry Ireys, Melissa Azur, and Cara Stepanczuk
DATE: 4/9/2019
SUBJECT: Request for reconsideration of candidate measures #3443 and #3445

Candidate Measure #3443: All-cause emergency department utilization rate for Medicaid beneficiaries with complex care needs and high costs (BCNs)

Candidate Measure #3445: All-cause inpatient admission rate for Medicaid beneficiaries with complex care needs and high costs (BCNs)

Measure Steward: Centers for Medicare & Medicaid Services (CMS), Centers for Medicaid & CHIP Services

Measure Developer: Mathematica Policy Research (Mathematica)

Review Cycle: All-Cause Admissions and Readmissions, fall 2018

CMS and Mathematica request a reconsideration of the National Quality Forum (NQF) All-Cause Admissions and Readmissions Standing Committee's decision to not recommend candidate measures #3443 and #3445 for endorsement. The Committee did not recommend endorsement because of concerns with the measures' validity. Specifically, the Committee believed that their validity was undermined by three factors:

1. Medicaid populations differ across states and, as a result, differential performance on these measures across states will be difficult to interpret because such differences could be due to underlying differences in the Medicaid populations or to actual differences in performance.

2. Data from 10 states do not provide a sufficiently representative sample of the national Medicaid population.

3. Variable quality of Medicaid data.

We request reconsideration of the two measures. As discussed below, we believe the measures meet NQF's validity criteria.

Differences in Medicaid populations across states. State Medicaid programs vary substantially, both with respect to the populations they cover and the quality of data they report to CMS. States differ because of a fundamental premise of the nation's Medicaid legislation: The Medicaid program is a federal-state partnership. Congress can alter the legislation and CMS can require states to take certain actions, but states retain substantial authority and responsibility for operating Medicaid programs. By design, major differences among state Medicaid programs are inherent. We believe that, in the case of these two measures, the undue emphasis on state variation in Medicaid programs moved the validity bar to an unrealistically high level.

We recognize that the Committee has to make decisions about validity based on the strength of evidence. However, due to data limitations in the Medicaid landscape, not all measures will have the same amount of evidence behind them at the time they are submitted for endorsement. In fact, the committee noted that "the developer assessed face validity, which met the testing requirement for a new measure and the risk adjustment model demonstrated adequate discrimination and calibration." In order to move the Medicaid measurement field forward, the Committee will need to continue to recommended, as it has done in the past, Medicaid measures for endorsement despite certain limitations.

Representative sample of Medicaid population. With respect to the second concern noted above, many or perhaps even most researchers who have experience with Medicaid programs would view our 10 states as providing a robust sample for assessing measure performance. They vary in location, geography, size, and delivery system (fee-for-service or managed care). Perhaps most importantly, they all provide high-quality data. Given that Medicaid programs differ with respect to eligibility policies, mix of delivery models, payment rates, and other features, any sample of programs can be challenged as not being representative of all 50 states. The goal of measure testing is not to represent the Medicaid program in general; rather, it is to select a diverse group of states that have high-quality data and whose populations capture, for the key variables in question, the majority of the variation that also occurs within other states. Given the diversity of the states used in our analyses, we believe our analysis achieves this goal.

Additionally, we designed the measure specifications to maximize the likelihood that states would define the relevant denominator population consistently. Specifically, the BCN population definition requires 10 months of Medicaid eligibility in the definition year; this enrollment threshold is as inclusive as possible while preserving enough data to determine whether one could be considered part of the BCN population. The BCN population definition requires at least 1 inpatient admission and at least 2 chronic condition diagnoses in a 12-month period; these criteria produce a population that matches how the BCN population is characterized in practice and the scientific literature; the population is older, more likely to have co-occurring health conditions, and more likely to be eligible for Medicaid based on aged/disabled status than the non-BCN population. The BCN measures are calculated on a beneficiary-month basis, which mitigates the effect of churn on measure validity. A Medicaid beneficiary that meets the BCN population definition in the lookback year can contribute to the measure regardless of the number of months enrolled in the measurement year.

In general, researchers seek a study sample that is representative of a population because they want to ensure that results can be generalized to that population. But no researcher or state program administrator assumes that measure-performance findings based on one state Medicaid program (or multiple states or the nation as a whole) can be applied to another state without substantial scrutiny of the underlying assumptions and without ensuring that appropriate data can be obtained. We believe that the Committee should take this standard practice of reviewing measure-performance findings based on a study sample into account in applying the criterion of representativeness to these two measures.

Concerns about data quality. In addition to the two primary concerns noted above, the Committee observed that states vary considerably with respect to quality of Medicaid data

reported to CMS. We worked to evaluate the quality of relevant data in all the states and selected those states whose data met our quality standards. Specifically, as described in the original measure submission, the states we chose for testing had indicators that aligned with national inpatient and emergency utilization benchmarks and did not have data anomalies that would raise analytic problems (such as high levels of missing data). Further, NQF has endorsed measures in the past that were tested with Medicaid data from the same data source.

It is very important to understand the nuances and differences in the quality of Medicaid data available to CMS. While CMS has made great strides in improving the types and quality of data collected by states and reported to CMS, this effort is still very nascent and will take a very long time to complete. We do not believe that state variation in data quality should be a key factor in determining suitability of Medicaid measures for endorsement as long as the data used for development is of sufficient quality. As noted above, this measure was developed using a diverse group of states that have high-quality data. We also note that even perfect data from all states will not change the fact that state Medicaid programs have substantially different design and operational features.

We hope that the Committee will accept our request to reconsider these measures in light of the issues we have raised in this memo.

cc: Roxanne Dupert-Frank, CMS

Helen Dollar-Maples, CMS

Amy Gentile, CMS