NATIONAL QUALITY FORUM

Memo

Date: October 22, 2015

To: Cost and Resource Use Standing Committee &

CMS/Yale Measure Development Team

From: NQF Cost and Resource Use Project Team

Re: Review of SDS Empirical Analysis for CMS/Yale Cost Measures

The Cost & Resource Use Standing Committee will meet via webinar on Tuesday, October 27.

The purpose of the meeting is to:

- Provide an overview of the process and plan for reviewing the (3) CMS/Yale cost measures for cardiovascular and pneumonia conditions under the new guidance for sociodemographic status (SDS) risk adjustment.
- Review and discuss the empirical analysis of the selected SDS risk adjustment factors for the (3) cost measures
- Prepare the Committee to make final recommendations on the validity and endorsement status for the measures under review.

Standing Committee Action:

1. Review the Yale memo discussing responses to prior recommendations and empirical analysis of SDS risk factors.
2. Review this memo; prepare to provide input and discuss the Committee discussion questions on page 6.
3. Submit vote on validity criterion and endorsement recommendation.

Agenda

3:00pm Welcome & Roll Call
3:05pm Background
- How did we get here?
- Goals and purpose of this call

3:15pm Review of Empirical Analysis
- Developer overview and summary of memo and findings
- Committee Discussion

4:45pm Public and Member Comment
4:55pm Next Steps
5:00pm Adjourn

Overview: Reviewing the Cost Measures during the SDS Trial Period

The SDS trial period approved by the NQF Board of Directors is designated as a 2-year period of time during which SDS factors should be considered for potential inclusion in the risk-adjustment approaches if there is a conceptual reason for doing so. If there is a conceptual relationship between potential SDS risk factors and the outcome of interest, the developer

should conduct empirical analyses to determine whether such factors should be included in the risk-adjustment approach.

Following the NQF Board of Directors Executive Committee decision to endorse the cost measures with the condition that they be considered under the trial period guidance, NQF, in collaboration with the CMS/Yale measure development team, agreed to divide the assessment of the impact of SDS variables into two stages (and webinars).

Measures under Review

- #2431: Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction (AMI) (CMS/Yale)
- #2436: Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Heart Failure (HF) (CMS/Yale)
- #2579: Hospital-level, risk-standardized payment associated with a 30-day episode of care pneumonia (CMS/Yale)

Stage 1/Webinar #1 (May 21, 2-4pm ET): Conceptual Analysis

- Review conceptual analysis of relationships between SDS factors and the outcomes of interest
- Determine whether further empirical analysis is warranted
- Identify the variables to be pursued in empirical analysis
- Provide input on the plan or approach to empirical analysis of the selected variables

Stage 2/Webinar #2 (October 27, 3-5pm ET): Empirical Analysis

- Review empirical analysis of the impact of SDS risk factors in the risk model and measure scores:
    - Vote on Validity Criterion
- Make a recommendation on endorsement status:
    - Recommend [continued] endorsement of the measure OR
    - Recommend to de-endorse the measure

Conceptual Analyses Review

For the May webinar, the CMS/Yale Core development team submitted a memo summarizing their literature review and conceptual model diagram illustrating the potential relationships between various factors during the episode of care captured by the measures (i.e., hospital admission through 30 days post-discharge). Based on this conceptual analysis, they identified three variables that have been identified in the literature to have a conceptual relationship to utilization and payment.

Proposed Variables:

- Patient zip code (proxy for educational attainment or income)
- Medicaid status (proxy for low income and insurance coverage)
- Black or white race

Summary: Committee Discussion and Recommendations on Conceptual Analyses (Webinar 1)

1. **Broaden the conceptual model**. The Committee was concerned that the conceptual model seemed too medical-oriented and should be broadened to account for more public health variables. For example, the model did not address community, environmental, or patient factors (e.g., social supports, lack of money to buy medication, no refrigerator). The conceptual model should reflect resources available for care within individual hospitals. While these should not be included in the risk-adjustment approach, because differential resources can impact quality of care, they should be noted in the conceptual model.
2. **Additional literature review.** The Committee believed that further literature review was needed to determine the within and between effects of race on hospital performance. Some members strongly suggested that between and within hospital differences should be a lens through which this information should be analyzed. Members also suggested that the developers do a broader search of literature to include readmissions and impact of SDS on health.
3. **Conceptual Relationships.** Based on the research performed by the developers, the Committee agreed there is a conceptual relationship between the selected variables and cost/payment outcomes.

In response to the Committee's recommendations (1 &2 above) in May, the developers performed additional literature review and revised their conceptual model (Appendix 1 of Yale Memo). Text from this memo summarizing the key findings from this additional research has been excerpted below:

- They reviewed 14 additional articles that examined within and between hospital differences in outcomes related to SDS variables:
    - "Taken together these papers do not present a conclusive or consistent picture about the role of within hospital differences in treatment of patients based on SDS nor the subsequent impact on outcomes or cost. However they provide some evidence that in certain settings differential care by race could contribute to differences in costs and outcome." (page 3)
    - "Taken together, the body of literature reveals an inconsistent and complex association of low SDS and health outcomes. Most studies used race as their independent variable with less attention to income or other measures of poverty (e.g. Medicaid status). The literature demonstrates both within and between hospital differences in outcomes among racial/ethnic groups that can be partially explained by the use of lower quality hospitals by minorities." (page 3)

Standing Committee Discussion:

1. Does the committee believe the developer has adequately supplemented their conceptual analysis based on previous Committee recommendations?
2. Does the conceptual model adequately reflect the impact of SDS factors in the episode of care that is captured in these measures?

Empirical Analyses Review

NQF guidance for the evaluation of SDS factors states that if the Committee believes a conceptual relationship exists between the sociodemographic factor(s) and the outcome (i.e, resource utilization or cost), developers should conduct empirical analyses to confirm that relationship.

During the May webinar, the Committee determined there is conceptual relationship between the proposed variables and the three cost outcomes. Their discussion yielded the following recommendations regarding the examination and consideration of these variables in empirical analyses:

- **Race:** The Committee recommended that the Yale team review the data and consider including other race variables beyond black.
- **Income and educational attainment**: The Committee was not in favor of the developers beginning empirical analysis using 5-digit zip code data. The Committee would prefer for the developers to use their resources analyzing the 9-digit zip code data once it is available to them.
- **Medicaid/dual eligibility status**: The Committee was in support of empirical analysis on this (Medicaid status) variable, but only in combination with the Low Income Subsidy (LIS) data as proxy for insurance status and income.

Current NQF guidance for the submission of empirical analysis of SDS factors in the risk model requires the submission of:

- Analyses and interpretation resulting in decision to include or not include SDS factors in section.
- Performance scores and risk model performance of the model with and without SDS factors included (including method and results).
- An interpretation of their results in terms of the differences in performance scores for the same entities.
- Submission of updated reliability and validity testing and specifications for a stratified version of the measure using these factors, if SDS factors are included in the risk-adjustment approach.

The importance of the SDS variables in the risk adjustment model should be evaluated by the size of the SDS coefficients in the risk adjustment model, the p-values associated with the SDS coefficients, and the impact of adjusting for the SDS variables on the measure results. Reasons for including the SDS variables in the risk-adjustment approach include (1) demonstration of the contribution of the SDS factor(s) to unique variation in the outcome that is not due to between-unit effects, (2) adjustment leads to substantial differences between measure scores (although this doesn't have to result in change in rankings), or (3) if needed for face validity of the approach.

Summary: Key Findings Discussed in the Yale Empirical Analysis Memo

Variables Used in the Empirical Analysis:

1. Race: Categorized as Black and Non-Black
2. Medicaid enrollment/Dual Status (as proxy for low income): Categorized as Medicaid and Non-Medicaid.

Definitions (Table 1):

- Identified conceptual relationship: The Committee's determination of whether the variables had a conceptual relationship to the cost outcomes.
- Variation in prevalence of SDS factors across entities: If the prevalence of a particular factor does not vary across the measured entities, then adjustment likely is not necessary. But if it does vary substantially, then there is reason to believe that one should potentially control for it in the risk-adjustment approach.
- Bi-variate relationship between SDS factors and outcome: If the SDS factor is associated with the outcome, then it is a potential confounder and may be a candidate for risk adjustment.
- Significant to the multivariable model: Analysis demonstrating that the SDS factor is statistically associated with the outcome of interest, after controlling for other (clinical) factors
- Good model fit: The risk-adjustment model adequately "reflects" the data.
- Impact on the risk model and measure scores with the inclusion of the SDS Factors: How the measure scores and risk model fit is impacted when you add in the SDS factors.
    - Improvement in model fit with the addition of variables (based on Quasi R2)
    - Change in distribution of scores:
    - Change in mean Risk Standardized Payment (RSP):
    - Impact on Risk Standardized Payment:
- Decision to include variables in the model: The developer's decision on whether or not to include the variables in the risk model.

Table 1: Assessment of Empirical Analyses Results and Validity of the Risk Adjustment Approach (Applies to Black Race + Medicaid Variables)

Assessment of Validity of the Risk Model	AMI	HF	PN
Identified conceptual relationship	Yes	Yes	Yes
Variation in prevalence of SDS factors across entities	Yes (Tables 2 &3)	Yes (Tables 11 &12)	Yes (Table 20 &21)
Bi-variate relationship	Yes (Table 7)	Yes (Table 16)	Yes (Table 25)
Significant to the model	Yes (Table 8)	Yes (Table 17)	Yes (Table 26)

Assessment of Validity of the Risk Model	AMI	HF	PN
Good model fit	Yes (Table 9)	Yes (Table 18)	Yes (Table 27)
Impact on the risk model and measure scores with the inclusion of the SDS Factors:			
Improvement in model fit with the addition of variables (based on Quasi R^2)	Slight (Table 9)	Slight (Table 18)	None (Table 27)
Change in distribution of scores	Slight (Table 6)	Slight (Table 15)	Slight (Table 24)
Change in Mean Risk Standardized Payment (RSP)	None (Table 4)	$1 (Table 3)	$1 (Table 22)
Impact on RSP	Black: Lower payment: beta: -0.058) Medicaid: Lower payment (beta: -0.017) (Appendix 3)	Black: Lower payment: beta: -0.030) Medicaid: Higher payment (beta: 0.012) (Appendix 4)	Black: Higher payment: $287 Medicaid: Higher payment: $496 (Appendix 5)
Decision to include variables in the model	No	No	No

Standing Committee Discussion:

3. Does the committee believe the developer has adequately demonstrated validity of their risk adjustment approach?
4. Has the developer adequately supported their decision to not include the SDS variables in the risk model?

Standing Committee Action:

1. Submit your votes on validity and endorsement with rationale. ***(Please only submit your vote ON or AFTER the October 27 webinar)***
    a. Validity: High, Moderate, Low or insufficient
        i. Consider the other validity sub-criteria when submitting your votes (See *Appendix A* for Resource Use Measure Evaluation Criteria for Validity and *Appendix B* for prior voting results on validity).
    b. Recommendation for endorsement:
        i. Recommend for [continued] endorsement (without the inclusion of SDS) factors in the risk model
        ii. Do not recommend for [continued] endorsement

Appendix A: *Resource Use Measure Evaluation Criteria (Validity Only)*

2b. Validity

2b1. The measure specifications are consistent with the measure intent described under criterion 1c and captures the most inclusive target population.

2b2. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the cost of care or resources provided.

2b3. Exclusions are supported by the clinical evidence.

AND/OR

There is a rationale or analysis demonstrating that the measure results are sufficiently distorted due to the magnitude and/or frequency of the non-clinical exclusions;

AND

− Measure specifications for scoring include computing exclusions so that the effect on the measure is transparent (i.e., impact clearly delineated, such as number of cases excluded, exclusion rates by type of exclusion);

AND

− If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b4. For resource use measures and other measures when indicated:

− an evidence-based risk-adjustment strategy (e.g., risk models, risk-stratification) is specified and is based on patient clinical factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care and has demonstrated adequate discrimination and calibration OR

− rationale/data support no risk-adjustment/-stratification.

2b5. Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful13 differences in performance.

2b6. If multiple data sources/methods are specified, there is demonstration that they produce comparable results.

2c. If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results (e.g., by race, ethnicity, socioeconomic status, gender) OR rationale/data justifies why stratification is not necessary or not feasible.

Appendix B: *Summary of Committee Deliberations of Scientific Acceptability (Initial Endorsement)*

2579 Hospital-level, risk-standardized payment associated with a 30-day episode of care for pneumonia

Description: This measure estimates hospital-level, risk-standardized payment for a pneumonia episode of care starting with inpatient admission to a short term acute-care facility and extending 30 days post-admission for Medicare fee-for-service (FFS) patients who are 65 years of age or older with a principal discharge diagnosis of pneumonia.
Resource Use Measure Type: Per episode
Level of Analysis: Facility
Costing Method: Standardized pricing
Target population: Senior Care
Data Source: Administrative claims
Measure Steward: Centers for Medicare & Medicaid Services
STANDING COMMITTEE MEETING [06/25/2014]

---

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria
(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)
2a. Reliability: **H-10**; **M-11**; **L-1**; **I-0** 2b. Validity: **H-3**; **M-18**; **L-1**; **I-0**
Rationale:
• The Committee stated that the measure specifications were precise and that the measure was well-constructed. This measure captures risk-standardized payments for a thirty-day episode of care for Medicare patients diagnosed admitted to the hospital with a diagnosis of pneumonia through administrative claims data.
• The developer provided reliability testing at the level of the performance measure score; testing was performed by calculating the Intraclass Correlation Coefficient (ICC) score by calculating the risk standardized payment using a split-sample of the combined 2008-2009 data from hospitals. The ICC score was 0.825, indicating significant agreement between the two samples, which the Committee found sufficient.
• The Committee questioned the validity of specifying the measure for a thirty-day episode triggered by admission for pneumonia, as the treatment of pneumonia may require care coordination post-discharge that may extend past thirty days. The Committee stated that this could affect payments captured during the post-discharge period, artificially inflating or deflating the costs for some patients simply because of the construct of the measure.
• The Committee raised concerns regarding the attribution approach and the implications for attribution of costs if a patient were transferred to another hospital. The developer clarified that only 0.4 percent of cohorts are transferred for pneumonia, which represents a small number of beneficiaries.

In the case of transfer patients, costs for the patient will be attributed to the initial admitting hospital, as hospitals are increasingly responsible for care delivered up to 30 days after discharge. The Committee found this approach to attribution to be acceptable.
• The Committee stated concern that the low r-squared value (.07) for the risk model may indicate that case mix is not being appropriately adjusted for through the risk model. The developer clarified that at lower patient volumes, there is less certainty when estimating cost. The measure uses a continuous outcome which results in a more accurate estimate than would result from a binary outcome. Additionally, the measure uses hierarchical risk modeling that adjusts hospitals with low patient volume towards the mean. The Committee found this explanation to be sufficient.
• The Committee questioned whether adjustments for sociodemographic status (SDS) factors should be incorporated into the risk adjustment model. NQF clarified that it is in the early stages of reviewing our policy on risk adjusting for SDS factors. The recommendations for modifying NQF's current policy on adjusting for SDS factors have not yet been finalized. As such, we ask that Committees continue to evaluate measures according to our current guidelines, that SDS factors are not included in the risk adjustment model, but are used to stratify the measure. If in the future the recommendations for adjusting for SDS factors become NQF policy, measures that may be improved from incorporating these adjustments will be updated and reviewed by the Committee through one of NQF's measure maintenance processes.

2431 Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction (AMI)*

Submission | **Specifications**

Description: This measure estimates hospital-level, risk-standardized payment for an AMI episode-of-care starting with inpatient admission to a short term acute-care facility and extending 30 days post-admission for Medicare fee-for-service (FFS) patients who are 65 years of age or older with a principal discharge diagnosis of AMI.
Resource Use Measure Type: Per episode
Level of Analysis: Facility
Costing Method: Standardized pricing
Target Population: Senior Care
Data Source: Administrative Claims
Measure Steward: Centers for Medicare and Medicaid Services
STANDING COMMITTEE MEETING [March 4-5, 2014]

---

2. Scientific Acceptability of Measure Properties
(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)
2a. Reliability: **H-3; M-16; L-2; I-0** 2b. Validity: **H-0; M-9; L-7; I-4**
Rationale:
• The Committee raised concern about the ability to assess performance of low volume hospitals given the hierarchical modeling approach and the potential implications it could pose for the reliability and validity of the measure. The developers responded that at lower patient volumes, the less certainty you have about your estimates for cost. This measure uses a continuous outcome so the estimate is more accurate than a binary outcome.

Additionally, this measure uses hierarchical risk modeling that adjusts hospitals with low patient volume towards the mean. Furthermore, reporting is only done for hospitals that have 25 or more cases.

• The Committee further questioned the decision to attribute the entire cost of an episode to the initial hospital in the case of a transfer to another facility. The developers responded that the decision was made not to exclude these cases because transfers account for approximately 8 percent of AMI episodes. This represented too many cases to exclude. Furthermore, the initial hospital begins the episode of care and can have a great influence over the coordination of care.

• The Committee raised concerns about whether the supplied reliability testing was done with the amount of data required by the specification of the measure. The measure is specified for a 12-month period and the testing used combined 2008 and 2009 data. The developers responded that the measure will eventually be implemented with three years of data but when the testing was performed, only two years of data was available. The decision to include three years of data was made to include as many hospitals in the measurement as possible. Many hospitals do not have 25 AMI cases in a year and would therefore not meet the threshold for reporting.

• In addition to the risk adjustment provided in the overarching issues section, the Committee was concerned that the developer did not do empiric measure-level validity testing for the measure as specified. The developers acknowledged that they relied on prior research on risk adjustment testing for mortality measures and also relied on face validity testing with their technical expert panel.

2436 Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for heart failure (HF)

Description: This measure estimates hospital-level, risk-standardized payment for a HF episode of care starting with inpatient admission to a short term acute-care facility and extending 30 days post-admission for Medicare fee-for-service (FFS) patients who are 65 years of age or older with a principal discharge diagnosis of HF.

Resource Use Measure Type: Per episode

Level of Analysis: Facility

Costing Method: Standardized pricing

Target Population: Senior Care

Data Source: Administrative Claims

Measure Steward: Centers for Medicare and Medicaid Services

STANDING COMMITTEE MEETING [March 4-5, 2014]

---

2. Scientific Acceptability of Measure Properties

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-7; M-11; L-2; I-1** 2b. Validity: **H-0; M-9; L-6; I-5**

Rationale:

• The Committee questioned the description of a “typical heart failure” patient considering that many patients have chronic heart failure and a hospitalization occurs for an acute incidence of the disease.

The developer responded that they meant non-LVAD, non-transplant, non-major surgical procedure heart failure patients. These conditions dramatically change the payment outcome. They are sicker patients and were excluded from the measure.

• The Committee also questioned the methodology for choosing the index admission for patients who might have multiple hospitalizations in the same year for heart failure. The developer responded that the hospitalization is randomly selected and any re-hospitalization within 30 days of that index admission would be considered a re-admission and counted in the total hospitalization cost.

• The Committee expressed concern that attributing costs to hospitals was inappropriate for heart failure patients and that the real accountability should be with the ambulatory providers. Furthermore, the 30-day time period for costs does not align with the typical disease progression for a heart failure patient. A longer period, perhaps 12 months, would be more appropriate for the chronic nature of this disease.

• The developer defended the attribution to the hospital by stating that heart failure is a leading cause of hospitalization for the elderly and it represented a high-leverage opportunity to measure and evaluate spending. Additionally, the 30-day time period was short enough that the associated spending would be attributable to the hospital admission.

• In addition to the risk adjustment discussion provided in the overarching issues section, the Committee was concerned that the developer did not do empiric measure-level validity testing for the measure as specified. The developers acknowledged that they relied on prior research on risk adjustment testing for mortality measures and also relied on face validity testing with their technical expert panel.
