Center for Outcomes Research and Evaluation (CORE)
1 Church Street, Suite 200
New Haven, Connecticut 06510-3330
Phone: 203-764-5700 Fax: 203-764-5653

# MEMORANDUM

**TO:** National Quality Forum (NQF) Surgery Standing Committee
**FROM:** Mayur Desai, PhD, MPH, Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (YNHHSC/CORE)
**THROUGH:** The Centers for Medicare & Medicaid Services (CMS)
Vinitha Meyyur, PhD
**DATE:** Friday, June 19, 2015
**SUBJECT**: Response to post-evaluation public comments regarding NQF 2687: Hospital Visits after Hospital Outpatient Surgery; Submitted June 19, 2015

## 1. American Hospital Association (AHA) Comment and Response

**AHA comment (Part I):**
The American Hospital Association (AHA) strongly urges the National Quality Forum (NQF) Surgery Committee not to support the endorsement of NQF # 2687, Hospital Visits after Outpatient Surgery, until it has fully examined the impact of sociodemographic factors on performance. We are deeply concerned that at a time when the NQF is undertaking a "trial period" to assess the impact of sociodemographic factors on outcome measures, it is not clear the committee was even asked to consider the impact of such factors. This lack of clarity is especially perplexing since it appears the measure developer actually provides some information on the potential the impact of two sociodemographic factors – race and dual-eligibility for Medicare and Medicaid – in the measure submission. The developer's analysis seeks to assess the impact of using patient-level adjustment for dual eligibility and race. The analysis also examines the distribution of hospital outpatient department (HOPD) performance scores by the proportion of African-American and dual-eligible patients they treat. The developer asserts that sociodemographic adjustment is unnecessary because the measure scores of adjusted and unadjusted measures using the patient level adjustment are highly correlated, and because there is "substantial overlap" of scores for HOPDs with lower and higher proportions of dual-eligible patients.

Yet, there is no indication in the draft report that this analysis was discussed by the committee. It is also not clear whether the committee would even have criteria from NQF to judge the validity of the analysis. We believe a deeper examination of sociodemographic adjustment – based on clearly articulated criteria from the NQF – is needed, especially since the developer's

analysis shows a number of high outliers among HOPDs with the largest proportion of dual-eligible patients. The NQF's 2014 expert panel report on sociodemographic adjustment offers a useful framework for potential evaluative criteria, and underscores the need to expand the analysis. For example, it is unclear how the developer could reach such a sweeping conclusion about the necessity of sociodemographic adjustment using only race and dual-eligible status. Other factors – such as income or whether the patient lives alone, or access to pharmacy care – may have tighter conceptual links to the likelihood of needing hospitalization after surgery.

**Response:**
We appreciate your concern about the potential effects of SDS on the measure score. We wanted to address your comments on both the process of review and the substance of our conclusions in the NQF application based on the SDS analysis we conducted for the application.

Regarding the process, the surgery measure is not technically in NQF's SDS pilot.
"This trial period went into effect on April 15, 2015. This means that projects with measure submission deadlines before that date fell under NQF's previous policy/guidance on SDS adjustment, while projects with measure submission deadlines after that date are subject to the trial policy on SDS adjustment. Since the 2015 Surgery project's measure submission deadline was January 14, 2015, both the developer and the Surgery Standing Committee conformed to the [pre-trial] policy regarding inclusion of SDS factors in the risk-adjustment approach (email from Andrew Lyzenga at NQF, June 15, 2015)."

Regarding the substance of your concern, consistent with the pre-trial NQF guidance on SDS, we evaluated the potential effects of risk adjusting for two SDS indicators – Medicaid-dual eligibility and race. These variable are readily available in the CMS claims data. In addition, use of Medicaid eligibility status as a proxy for SDS is consistent with prior research as well as NQF recommendations
(http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx). Our results show that adjusting for these factors at the patient level does little to change the measure scores; unadjusted and adjusted HOPD risk-standardized hospital visit (RSHV) ratios are highly correlated (Pearson correlation 0.990 and 0.998 for adjustment for Medicaid-dual eligibility and race, respectively). This suggests that including a patient-level risk adjuster for SDS will make little difference in the measure results after accounting for other factors already adjusted for in the model, such as age, comorbidities, and the complexity of the surgery.

In addition, to explore whether there might be differences in HOPD RSHV ratios by the proportion of lower SDS patients hospitals care for, we examined the distribution of measure scores by quartiles of both percentage of dual-eligible patients and percentage of African American patients. Although the results show a trend toward higher measure scores in the highest quartile of lower SDS patients, they also show that some hospitals with relatively high proportions of lower SDS patients can and do perform well on the measure. We cannot tell from these analyses what is causing the observed differences across quartiles of proportion of lower SDS patients. One of the potential causes is differences related to quality. For example, some hospitals may be better able than other hospitals to meet the needs of patients with low

literacy. Given these findings, on balance we do not recommend adjusting the measure for SDS at this time. Doing so will not appreciably change the measure scores and might contribute to masking disparities in care.

CMS is participating fully in the NQF trial and is actively working to further consider issues related to adjusting for SDS. In addition, CMS notes that the Office of the Assistant Secretary for Planning and Evaluation (ASPE) is conducting research on the issue of risk adjustment for socioeconomic status as directed by the IMPACT Act and will issue a report to Congress by October 2016. CMS will closely examine the recommendations issued by ASPE and consider how they apply to this and other CMS quality measures.

**AHA comment (Part II):**
The AHA and numerous other stakeholders have repeatedly noted that sociodemographic adjustment holds promise to improve the scientific acceptability – and therefore, credibility – of outcome measures whose performance can be influenced by factors beyond a provider's control. We would be especially interested to see a reliability analysis of a sociodemographically-adjusted version of NQF #2687 since the reliability of the measure without adjustment is disappointing. Indeed, the intraclass correlation coefficient for the measure shows only a "moderate" level of reliability. Given that this measure may be used in future accountability programs for HOPDs, we believe the public and hospitals require and deserve measures with more than "moderate" levels of reliability.

**Response:**
CMS did consider the effect of adjusting for SDS and reported the results in the NQF application. As discussed in the application and in response to the question above, we do not recommend adjusting for SDS at this time, so testing the reliability of the measure with SDS adjustment is not necessary at this time. As you note, reliability testing yielded an intraclass correlation coefficient (ICC) of 0.50, which according to conventional interpretation is "moderate." It should be noted that this ICC value is consistent with those of other CMS claims-based measures. In addition, measure testing was conducted using a 20% sample of Medicare Fee-for-Service data. We expect the reliability score will be higher in the national 100% sample where individual facility volumes would be higher yielding more reliable individual facility results. The 100% sample would be used for public reporting.

**AHA comment (Part III):**
The AHA also does not understand the committee's decision to defer an assessment of the need harmonize this measure with NQF # 2539 (Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy) until the next measure maintenance cycle. Harmonization is intended to ensure that measures on related topics use measurement approaches that are as consistent as possible. In light of the recent Institute of Medicine Vital Signs report showing how the lack of consistency in measures has impeded our nation's ability to measure and improve quality, we believe the NQF should use every available opportunity to promote harmonization of the measures it endorses.

**Response:**
The present measure (NQF # 2687) is already fully harmonized with NQF # 2539 (Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy) on areas of the methodology that are analogous. Specifically, both measures use the same outcome. For both the outpatient surgery measure and the outpatient colonoscopy measure, the outcome is identically specified as all-cause, unplanned hospital visits, defined as 1) an inpatient admission directly after the procedure, or 2) an unplanned hospital visit (emergency department visit, observation stay, or unplanned inpatient admission) occurring after discharge and within 7 days of the outpatient procedure.

**AHA comment (Part IV):**
Until this measure has been further assessed for the impact of sociodemographic factors, demonstrates better reliability, and has been appropriately harmonized with related measures, the AHA cannot support its endorsement as a national standard. Measures should not carry the imprimatur of NQF endorsement if they create confusion, limit our ability to identify real opportunities for improvement and lead to unfair and damaging judgments of provider performance.

**Response:**
Please see responses above addressing sociodemographic factors, reliability, and harmonization.

## 2. Federation of American Hospitals (FAH) Comment and Response

**FAH Comment (Part I):**
The Federation of American Hospitals appreciates the opportunity to comment on measure 2687, Hospital Visits After Hospital Outpatient Surgery. The FAH does not support endorsement of this measure for the following reasons.

The Committee report recommends measure 2687, however the Committee's review does not provide assurance that the measure has strong evidence of importance and indicates rather luke-warm reliability assessment. Specifically, the Committee reports states "…minimal evidence that ties specific processes to the outcome but that the rationale is sufficient to support the measure." The "rationale" is not scientific evidence of importance nor reliability and validity. It leaves readers wondering if the cost of implementing and using this measure, if it is endorsed, would yield any information that could help a facility improve care to its patients.

**Response:**
Thank you for the comment.

1) We believe that the measure will yield important information that will help facilities improve patient care. Measure testing demonstrated significant variation in risk-

standardized performance across facilities, indicating opportunities for quality improvement. Facilities with a higher than expected number of outcomes will be able to review and improve their processes around preparing the patient for surgery, the surgery itself, and follow-up care. In addition, in implementing the measure, CMS would provide each facility with patient-level data so that facilities could examine the specific causes of higher than expected outcome.

2) Reliability testing yielded an intraclass correlation coefficient (ICC) of 0.50, which according to conventional interpretation is "moderate." It should be noted that this ICC value is consistent with those of other CMS claims-based measures. In addition, measure testing was conducted using a 20% sample of Medicare Fee-for-Service data. We expect the reliability score will be higher in the national 100% sample where individual facility volumes would be higher yielding more reliable individual facility results. The 100% sample would be used for public reporting.

**FAH Comment (Part II):**
In addition, we are disappointed to see yet one more hospital revisit measure, similar in structure to a number of previous revisit measures, come through the NQF endorsement process without socio-demographic adjustment. This is particularly troubling since the measure 2687 is being evaluated during the SDS trial period and the measure has not been included in that trial process.

**Response:**
We appreciate your concern about the potential effects of SDS on the measure score. We wanted to address your comments on both the process of review and the substance of our conclusions in the NQF application based on the SDS analysis we conducted for the application.

Regarding the process, the surgery measure is not technically in NQF's SDS pilot.
"This trial period went into effect on April 15, 2015. This means that projects with measure submission deadlines before that date fell under NQF's previous policy/guidance on SDS adjustment, while projects with measure submission deadlines after that date are subject to the trial policy on SDS adjustment. Since the 2015 Surgery project's measure submission deadline was January 14, 2015, both the developer and the Surgery Standing Committee conformed to the [pre-trial] policy regarding inclusion of SDS factors in the risk-adjustment approach (email from Andrew Lyzenga at NQF, June 15, 2015)."

Regarding the substance of your concern, consistent with the pre-trial NQF guidance on SDS, we evaluated the potential effects of risk adjusting for two SDS indicators – Medicaid-dual eligibility and race. These variable are readily available in the CMS claims data. In addition, use of Medicaid eligibility status as a proxy for SDS is consistent with prior research as well as NQF recommendations
(http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx). Our results show that adjusting for these factors at the patient level does little to change the measure scores; unadjusted and adjusted HOPD risk-standardized hospital visit (RSHV) ratios are highly correlated (Pearson correlation 0.990 and 0.998 for adjustment for Medicaid-dual

eligibility and race, respectively). This suggests that including a patient-level risk adjuster for SDS will make little difference in the measure results after accounting for other factors already adjusted for in the model, such as age, comorbidities, and the complexity of the surgery.

In addition, to explore whether there might be differences in HOPD RSHV ratios by the proportion of lower SDS patients hospitals care for, we examined the distribution of measure scores by quartiles of both percentage of dual-eligible patients and percentage of African American patients. Although the results show a trend toward higher measure scores in the highest quartile of lower SDS patients, they also show that some hospitals with relatively high proportions of lower SDS patients can and do perform well on the measure. We cannot tell from these analyses what is causing the observed differences across quartiles of proportion of lower SDS patients. One of the potential causes is differences related to quality. For example, some hospitals may be better able than other hospitals to meet the needs of patients with low literacy. Given these findings, on balance we do not recommend adjusting the measure for SDS at this time. Doing so will not appreciably change the measure scores and might contribute to masking disparities in care.

CMS is participating fully in the NQF trial and is actively working to further consider issues related to adjusting for SDS. In addition, CMS notes that the Office of the Assistant Secretary for Planning and Evaluation (ASPE) is conducting research on the issue of risk adjustment for socioeconomic status as directed by the IMPACT Act and will issue a report to Congress by October 2016. CMS will closely examine the recommendations issued by ASPE and consider how they apply to this and other CMS quality measures.

**FAH Comment (Part III):**
Also, the committee points out that this measure should be harmonized with measure 2539 Colonoscopy Hospital Revisit Rate, but that harmonization won't happen until the next maintenance cycle. Such proposals were more understandable when NQF did not have standing committees in place. The FAH does not understand why harmonization and the SDS trial evaluation cannot be taken up at this point in time-- prior to endorsement.

**Response:**
The present measure (NQF # 2687) is already fully harmonized with NQF # 2539 (Facility 7-Day Risk-Standardized Hospital Visit Rate after Outpatient Colonoscopy) on areas of the methodology that are analogous. Specifically, both measures use the same outcome. For both the outpatient surgery measure and the outpatient colonoscopy measure, the outcome is identically specified as all-cause, unplanned hospital visits, defined as 1) an inpatient admission directly after the procedure, or 2) an unplanned hospital visit (emergency department visit, observation stay, or unplanned inpatient admission) occurring after discharge and within 7 days of the outpatient procedure.

**FAH Comment (Part IV):**
The continued proliferation of measures that are less than robust is very troubling. The FAH strongly urges NQF and the Committee to reconsider measure 2687 and to take this opportunity to change processes and procedures to do the appropriate SDS trail evaluation and

to harmonize measure 2687 with 2539. We cannot continue to push forward mediocre quality measures which do not provide solid, robust information to drive quality improvement and help hospitals meet the demands of the ever increasing pay-for-reporting programs. Newly endorsed quality measures should meet the standards to provide strong, reliable quality information on which hospitals and providers can act. The FAH encourages NQF not to give a stamp of approval to this measure that will not be well-received by the entities being measured and one that will not help to drive change in the delivery of care.

**Response:**
Responses to specific comments provided above.