TO: Consensus Standards Approval Committee (CSAC)

FR: Ashlie Wilbon, Lindsey Tighe, and Taroon Amin

RE: Cost and Resource Use Phase 2 for Cardiovascular Condition-Specific Measures: Member Voting Results

DA: July 2, 2014

The CSAC will review recommendations from the *Cost and Resource Use Phase 2 Cardiovascular* project at its July 9th in-person meeting.

This memo includes a summary of the project, recommended measures, public and member comment themes and their responses. .

This project followed the National Quality Forum's (NQF) version 1.9 of the Consensus Development Process (CDP). Member voting on these recommended measures ended on July 2nd.

Accompanying this memo are the following documents:

1. **Cost and Resource Use Phase 2 Cardiovascular Draft Report.** The draft report has been updated to reflect the changes made following Standing Committee discussion of public and member comments. The complete draft report and supplemental materials are available on the project page.
2. **Comment table.** Staff has identified themes among the comments received. This table lists 39 comments received and the NQF/Standing Committee responses.

Voting for Cost and Resource Use Phase 2 Cardiovascular measures closes at 6pm on July 2nd. Voting results will follow this memo as an addendum when available.

CSAC ACTION REQUIRED
Pursuant to the CDP, the CSAC may consider approval of 3 candidate consensus standards.

Cost and Resource Use Phase 2 Cardiovascular Measures Recommended for Endorsement:

- 1558: Relative Resource Use for People with Cardiovascular Conditions (NCQA)
- 2431: Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction (AMI) (CMS)
- 2436: Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for heart failure (HF) (CMS)

BACKGROUND
The draft report and comments received relate to the review of measures in the second phase of a three-phase effort to evaluate and endorse cost and resource use measures. In the first phase, a non-condition specific measure of total cost using a per-hospitalization episode approach for the Medicare population was endorsed. The Phase I report can be accessed via the NQF website. This second phase

focused on cardiovascular condition-specific measures and builds upon the phase one evaluation efforts and report.

The 25-person Cost and Resource Use Standing Committee reviewed three measures; NQF #1558 was recommended for endorsement by the NQF Cost and Resource Use Standing Committee; however, initially during the in-person meeting the Committee did not reach consensus for both NQF #2431 and NQF #2436. After consideration of all comments received during the NQF Member and public comment period, and additional justification for the measurement methodology and approach provided by the developer during the June 4th Committee call, the Committee requested the option to re-vote on their endorsement recommendation for both measures. The Committee subsequently voted to recommend both measures for endorsement. Pursuant to CDP process guidance, all three measures were posted for NQF member vote.

DRAFT REPORT

The Cost and Resource Use Phase 2 Cardiovascular Draft Report presents the results of the evaluation of three measures considered under the CDP. All three are recommended for endorsement as voluntary consensus standards suitable for accountability and quality improvement. The measures were evaluated against the NQF Resource Use Measure Evaluation Criteria.

	MAINTENANCE	NEW	TOTAL
Measures considered	1	2	3
Recommended	1	2	3
Not recommended	0	0	0
Reasons not Recommended	Importance- N/A Scientific Acceptability- N/A Overall- N/A Competing Measure- N/A	Importance- N/A Scientific Acceptability- N/A Overall- N/A Competing Measure- N/A	

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 prior to the evaluation of the measures via an online tool located on the project webpage. Third, NQF opens a 30-day comment period to both members and the public after measures have been evaluated by the full committee and once a report of the proceedings has been drafted.

Pre-evaluation comments

The pre-evaluation comment period was open from January 14th through February 3rd for the 3 measures under review. A total of two pre-evaluation comments were received, pertaining to the newly proposed CMS AMI and Heart Failure measures, raising concern about attribution of costs for these episodes to hospitals. These pre-evaluation comments were provided to the Committee prior to their initial deliberations held during the workgroups calls.

Post-evaluation comments

The draft report went out for public and member comment April 21st to May 21st. During this commenting period, NQF received 39 comments from 9 member organizations

A table of comments submitted during the comment periods, with the responses to each comment and the actions taken by the Standing Committee and measure developers, is posted to the Cost and Resource Use project page under the Materials section.

Comment Themes and Committee Responses

Comments about specific measure specifications and/or rationale were forwarded to the developers, who were invited to respond.

At its review of all comments, the Standing Committee had the benefit of developer responses. Committee members focused their discussion on measures or topic areas with the most significant and recurring issues.

Theme 1 – Appropriateness of attribution approach

Several commenters raised concern that, for the AMI (#2431) and Heart Failure (#2436) measures in particular, it may be inappropriate to attribute the cost of the episode to the hospital as much of the care happens in an outpatient setting. The proposed HF and AMI episode-of-care measures reflect the care of multiple providers across the health care delivery system. Like other measures that reflect the actions of many, attributing the results solely to one part of the system (i.e., hospitals) may not be appropriate. Commenters stated that measures should assess processes and outcomes over which the measured entity (e.g., hospital, physician group) can exercise a reasonable level of control, and that these measures may be more appropriate for an organization accepting bundled payments on behalf of all measured entities.

Committee Response:

The Committee acknowledged this concern; however, the Committee also stated that hospitals are increasingly responsible for care delivered up to 30 days after discharge. Consequently, hospitals are in the unique position of being able to push coordination of care, and this measure may serve as an impetus for this to occur.

Theme 2 – Adequacy of risk adjustment model

Several commenters stated that the low r-squared values for the AMI (#2431) and Heart Failure(#2436) measures (0.05 for AMI, 0.03 for HF) indicated that the risk model did not account for enough of the variation in measure scores and may not adequately account for patient case mix and severity. Moreover, commenters believe that the low level of reliability demonstrated illustrated another fundamental flaw of both measures—that they fail to adequately account for complicating conditions that patients have prior to an episode of care. These complicating conditions can markedly affect the costs of treatment, but the current risk adjustment model has limited ability to distinguish between conditions that a patient already has, and those that develop during the episode. Without risk adjustment or exclusions, a measured entity could appear to have higher costs than others simply because it cares for more complex patients.

Committee Response:

The Committee agreed with the developer's response 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.

Theme 3 – Approach to addressing transfer patients
For the AMI (#2431) and Heart Failure(#2436) measures, several commenters stated concern that the initial admitting hospital would be attributed cost for the episode when transferring patients to a second hospital, as the initial admitting hospital may have little control over the care that happens after the transfer. Commenters suggested that the developer exclude transfer patients from the measure.

Committee Response:

Prior measures under consideration by this Committee have raised this issue, particularly endorsed measure NQF #2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB).

In this measure all beneficiaries who are transferred are excluded from the measure; the Committee discussed the appropriateness of this exclusion at length. The developers explained that during their public comment on the measure, community hospitals argued that they should not be responsible for patients whom they stabilize and transfer to another facility. Facilities that receive transfers argued that they should not be responsible for care that was provided prior to the patients entering their facility. To account for both perspectives, the developer chose to exclude all transfers from the measure. ***The Committee noted that hospitals are increasingly responsible for care delivered up to 30 days after discharge; thus they agreed that hospitals should be responsible for the utilization and associated costs for patients that they transfer to other facilities. The developer acknowledged that it was challenging to address the various perspectives on attribution of transfers but agreed to reconsider the specification based on the Committee's feedback.***

The Committee again acknowledged this concern for these measures and reiterated the upward trend of hospitals responsibility for care delivered up to 30 days after discharge.

Theme 4 – Risk adjustment for socio-demographic factors
Several commenters stated that the risk adjustment models for the measures should capture socio-demographic factors, as there is robust evidence that such factors affect health outcomes, including resource use. The commenters noted that there is an NQF draft report which proposes including socio-demographic variables in risk adjustment models for some outcome and process measures.

NQF Response:
With respect to concerns that socio-demographic factors should be included in the measures' methodology, NQF appreciates these comments and is in the early stages of reviewing our policy on risk adjusting for socio-demographic factors. The report referenced is a draft report that has recently been reviewed during an NQF member and public comment period; the recommendations have not yet been finalized. As such, we ask that Committees continue to evaluate measures according to our current guidelines, that measures not be adjusted for socio-demographic variables. If in the future the recommendations for adjusting for socio-demographic variables become NQF policy, measures needing this adjustment will be updated and reviewed by the Committee through measure maintenance.

Committee Response:

The Committee acknowledged that the timing of the NQF risk adjustment report is not ideal; however, given the current NQF policy on adjusting for sociodemographic variables, the Committee requested that a recommendation be issued with the measure that when reported, the results should be stratified by sociodemographic variables.

NQF MEMBER VOTING RESULTS

NQF Member Voting on the Phase 2 Cardiovascular Conditions Cost and Resource Use measures closes at 6pm on July 2nd. Voting results will be sent separately from this memo and will be forwarded to CSAC when available.

Measure Evaluation Summary Tables

LEGEND: Y = Yes; N = No; H = High; M = Moderate; L = Low; I = Insufficient

#1558 Relative Resource Use for People with Cardiovascular Conditions

Description: The risk-adjusted relative resource use by health plan members with specific cardiovascular conditions during the measurement year.

Resource Use Measure Type: Per capita (population- or patient-based)

Level of Analysis: Health Plan, Population : National, Population : Regional

Costing Method: Standardized pricing

Target Population: Populations at Risk

Data Source: Administrative Claims

Measure Steward: National Committee for Quality Assurance

STANDING COMMITTEE MEETING [March 4-5, 2014]

1. Importance to Measure and Report

(1a. High Priority, 1b. Opportunity for Improvement, 1c. Measure Intent)

1a. High Priority: **H-20; M-2; L-0; I-0** 1b. Opportunity for Improvement: **H-7; M-13; L-2; I-1** 1c. Measure Intent: **H-17; M-6; L-0; I-0** 1. Overall: **H-12; M-10; L-1; I-0**

Rationale:

- National and regional health plan data aggregated by the developer highlight the clinical and financial importance of Cardiovascular Disease (CVD). The direct and indirect costs of CVD have increased from $400 billion to $500 billion from 2006 to 2010. When resource use data is presented alongside HEDIS quality composite, consumers, employers and government programs have a greater perspective on overall health plan value.

- The Committee noted during their initial measure evaluation that data on variations in cost and disparities in resource use in managing CVD were not included in the measure submission. The Committee also noted in their initial measure evaluation that no data was provided after the point of initial endorsement, which would have been helpful in assessing performance. The Committee mentioned these two concerns during the meeting, but the developer did not address either issue in the discussion. The Committee did not pursue further discussion.

- Though the developer stated that the benefit of the measure would be to gain greater information on the value of health care services through linking Relative Resource Use (RRU) measures and quality measures, the Committee's comments in the initial measure evaluation indicated that they were not entirely in agreement with the stated benefit of the measure. The Committee noted that "value" could be a difficult concept to define. One Committee member expressed the concern that while higher quality is always better, lower resource use may not always be better, especially when considering disparities in care that may result from undertreating particular groups. Another Committee member stated that the linkage between Relative Resource Use (RRU) measures and quality measures could be potentially useful for evaluation of benefits and programs.

    - The developer responded to the Committee's concerns by describing how the measure had been implemented successfully to provide greater information to plans. The Committee accepted the

developer's explanation and did not continue the discussion on measure intent.

2. Scientific Acceptability of Measure Properties

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

2a. Reliability: **H-0; M-18; L-2; I-3** 2b. Validity: **H-0; M-17; L-1; I-5**

Rationale:

- Both the Committee and the TEP expressed concerns regarding the reliability testing for the measure.

  Based on the measure submission, the Committee was not satisfied that reliability testing had taken place. Information submitted on reliability testing by the developer was descriptive, explaining that health plan stability was determined by the magnitude of quartile shifts of O/E over time, but no data on stability or magnitude was included in the measure submission. The developer responded to the Committee by explaining that plan data is tested annually and focused on the identification of outliers, errors in submissions and variations when correlating the measure data with other sources. The developer further explained that data on the percentage of health plans that had shifted quartiles could be found in the developer's annual report on RRU, but not the magnitude of shift for each health plan. The developer agreed to make portions of the annual report that discussed plan testing and quartile shifts in performance available to the Committee. The Committee is willing to accept additional data provided by the developer to support reliability; NQF staff will work with the developer to provide this information to the Committee after the NQF member and public comment period.

- One Committee member questioned if there was a minimum number of plan members needed per condition for the measure to be meaningful. The developer responded that the minimum number in the eligible population was 250 members, and that the risk adjustment had been validated against that number which should satisfy concerns regarding the small sample size. The Committee accepted the developer's explanation of risk adjustment validation.

- The Committee noted that the testing portion of the measure submission is primarily descriptive and indicates validity testing has not been performed. As a determination of face validity is adequate in evaluating resource measures, the Committee and the TEP asked the developer to discuss this in greater detail. The developer explained that the measure had been implemented in the marketplace and yearly performance analysis led to changes in the risk adjustment to the HCC model, a cap on the maximum amount of spending, additional exclusions, and lowering the number of members required for each plan. The Committee was willing to accept the developer's explanation as the measure has been implemented successfully in program use as proof of construct validity.

- One Committee member asked for clarification on the method of risk adjustment using gender as part of the model, and why that was chosen over the NQF preferred model of stratification. The developer explained that the HCC risk adjustment model requires gender as an input to predict utilization, but the data is reported back to health plans in a stratified fashion by risk cohort. The Committee accepted the developer's explanation of risk adjustment.

- The Committee also questioned the use of the RRU-HCC risk model as opposed to the CMS-HCC risk model in terms of the included comorbidities. The TEP noted the measure metrics were insufficient due

to the lack of reliability and validity testing results provided by the developer. The TEP could not identify the r^2 in the measure materials and agreed with the Committee that the risk model was not validated. The developer acknowledged the concerns of the TEP and stated that the original measure submission included validation information of the RRU-HCC model for applicability and appropriateness for the Relative Resource Utilization. The developer explained they are unable to produce an r2 on the aggregate data submitted by the plans as no individual patient data is provided. To prove the suitability of the RRU-HCC model, the developer tested the model using simulations of patient level data. When the developer was satisfied that the RRU-HCC model was valid, they applied that model to all health plan data used in development of the measure. The Committee is willing to accept the developer's explanation to determine face validity.

- The Committee agreed with the TEP concern that the exclusion of cardiovascular patients with HIV or cancer from the clinically relevant measure population was inappropriate. These patients use resources relevant to cardiovascular care, and the opinion of the TEP is their resource use should be captured. A health plan that refuses to pay for those resources could appear to perform better on this measure. The developer responded that these populations were excluded because of disproportionate resource use. The developer also stated that plans that refuse to pay for services were addressed through NCQA accreditation standards for each health plan, which include the process for approving or denying payment for services. The Committee was satisfied with the developer's explanation.

- The Committee questioned the impact on validity through exclusions created by instituting a maximum $100,000 spending cap per patient. The Committee's concern was that the spending cap artificially reduced variation by eliminating high dollar claims. The developer responded that the cost cap was developed based on modeling different RRU scenarios. These models were validated in 2005, 2009 and 2011. The Committee accepted the developer's explanation of the maximum spending cap.

- The Committee questioned the use of the HCC risk adjustment model. The Committee expressed the opinion that the use of this model would result in difficulty in determining between variations in resource use due to practice, and variations due to differences in patients. The developer explained that the HCC model of risk adjustment was chosen from four potential models as the HCC model demonstrated the best performance in terms of sensitivity and specificity to the measure population. Implementation of the HCC model increased the amount of data reported by health care plans on specific patient cohorts. The Committee accepted the developer's explanation of the choice of HCC risk model.

3. Feasibility: H-20; M-3; L-0; I-0

(3a. Byproduct of Care Processes; and 3b. Electronic sources; and 3c. Data Collection Strategy)

Rationale:

- The Committee was satisfied that the measure was feasible to implement as the measure is currently in use at both the health plan and the physician group level. The Committee acknowledged that the data is currently being collected and is available in electronic sources

4. Use and Usability: H-8; M-14; L-1; I-0

(4a. Accountability/transparency (used in accountability w/in 3 yr, public reporting w/in 6 yr, or if new - credible plan); and 4b. Improvement – progress demonstrated (if new - credible rationale); and 4c. Unintended

#1558 Relative Resource Use for People with Cardiovascular Conditions

Consequences - benefits outweigh evidence of unintended negative consequences (to patients/populations); and 4d. Measure Deconstruction – can be deconstructed to facilitate transparency and understanding)

Rationale:

- The Committee questioned how consumers and payers would use the measure for improvement with data reported at plan and population level. One member of the Committee emphasized that it was difficult to attribute meaning to changes in health plan performance as reported by the measure. The developer responded that they could offer guidance to users of the measure to identify cost opportunities. Measure results provide valuable information on patterns of utilization that are consistent with high quality and there are broad applications for this data for users. The intent is help the consumer and payer market understand the variation around cost and resource use. The Committee accepted the developer's explanation of how consumers and payers could use the measure.

- One Committee member noted that the measure submission sample report included information on planned use for regulatory and accreditation programs and asked the developer to comment on that application. The developer responded that inclusion into regulatory and accreditation programs is a potential planned use, dependent on the ability to more clearly differentiate between the performance of various health plans. The Committee accepted the developer's explanation.

- Another Committee member questioned if this measure would be integrated into the all payer claims database that a number of states are planning to implement, that allows for comparisons between plans. The developer responded that some of the participating states have limitations on the use of cost and resource data and that there were no immediate plans to integrate this measure into those programs. The Committee accepted the developer's explanation of restrictions on implementation in the all payer claims database.

- Some Committee members questioned the value of comparing variation across health plans and what actions health plans might take in result of these comparisons. These actions could have positive or negative implications. Actions could include network selection of providers, the implementation of value based purchasing programs, engagement with members or changes in medical policy that limit resources for patients with certain conditions. The developer responded that the pricing structure in the measure is standardized to eliminate market variation and all benchmarking and measure methodology is transparent to health plans, which allows for better comparison of quality between health plans and allows plans to examine their own performance to facilitate improvement. The developer stated that they did not expect that health plans would limit resources based results from this measure. Several Committee members indicated agreement with the developer by provided examples of use of this measure and other similar measures as in use as helpful feedback to delivery systems and as important in managing dual eligible populations. The Committee did not continue this discussion.

- One Committee member had a question about the application and use of this measure and questioned if the measure was is included in the Five Star Quality Rating system for comparison of RRU between Medicare Advantage Plans. The developer response was that that CMS had not included this measure in Five Star ratings. The Committee accepted the developer's explanation of measures included into Five Star rating.

#1558 Relative Resource Use for People with Cardiovascular Conditions

- One Committee member questioned how health plan performance could be determined by comparison between health plans. The developer explained that there are two indicators of health plan performance in annual data analysis. The first is a significant quadrant shift in health plans relative to each other; the second is an analysis of plan stability. These indicators allow for comparisons between health plans. The Committee accepted the developer's explanation of how comparison between plans could assist in understanding health plan performance.

- Some Committee members expressed the concern that the measure was specified to assess resource use at the health plan level and not the provider level. Those Committee members expressed the opinion that assessment at the provider level would provide more opportunity for improvement than the health plan level. The developer response was that health plan level assessment of resource use allows health plans to compare resource use to their peers, review their own data more closely, and look for opportunities and cost opportunities to improve based on the value they see from measure results. Individual plans can chose to apply assessments of resource use at the level of provider performance. The Committee accepted the developer's explanation of why plan level data was more useful than provider data in this measure.

- One Committee member expressed two concerns regarding the normalization of the data on health plan performance. The first concern questioned the value of normalization of data as it prevents the trending of information over time for a single health plan. The second concern questioned how normalization of data affects comparisons between health plans. The developer explained that in order to track improvement on the individual health plan level, a number of factors would have to be held artificially constant, which would prevent comparisons between health plans. All health plan submissions are combined with standardized prices and are used to calculate benchmarks. These standardized prices are updated yearly and the calculation of benchmarks is dependent on health plan submissions to the developer. Individual health plans can use this data to track their own improvement in different service categories, but the measure is constructed to allow for comparisons between health plans. The Committee accepted the developer's explanation of the reasoning behind data normalization.

Unintended Consequences

- The TEP and the Committee expressed concerns regarding the use of this measure for performance improvement. The TEP expressed the opinion that spending on cardiovascular conditions is not equitable for all populations, which results in disparities in care. Would improvement in performance be seen in reduced disparities in care? The TEP also questioned if all cardiovascular care considered under this measure was clinically effective or appropriate. If reduced spending by health plans indicates improvement in performance, how would that affect quality of care?

- The Committee expressed diverse opinions regarding the value of RRU of cardiovascular conditions for purchasers of health plans. One member expressed the concern that purchasers would focus more on plan cost than quality of plan and not consider health plan medical policy or provider network in their choices. Another member related the experience that health plans in local markets have used the relative

#1558 Relative Resource Use for People with Cardiovascular Conditions
resource index and pricing information from this measure and similar measures as feedback towards improvement. The developer explained that prior to the availability of this measure; purchasers of health plans only had information on cost, but not for quality of individual health plans. This measure allows purchasers of health plans, both states and employers, to compare quality and health plan performance between different plans.
5. Related and Competing Measures • No related or competing measures noted.
Standing Committee Recommendation for Endorsement: Y-20; N-2

6. Public and Member Comment: April 21, 2014 – May 21, 2014
• Several supportive comments for the measure were received, with commenters indicating that this measure could provide comparison data across the country. In addition, 2 issues were raised to the Committee's attention:
  ○ During the Phase 2 in-person meeting, the Committee requested that the developers provide a quantitative analysis of plan stability between measurement periods, including the magnitude and direction of shifts. The developers provided analysis demonstrating that a low proportion of plans change by more than one quartile.
  ○ A commenter raised concern that relative resource use measures are not particularly useful or meaningful to consumers to assess efficiency as they do not directly address out of pocket or total costs specific to the condition. The commenter requested that the Committee revisit the usability of this measure. The Committee and the developer acknowledged that this measure is less useful for patients and consumers; however, the Committee reaffirmed the importance of the measure for purchasers in particular that may use this measure to select a health plan.

7. Consensus Standards Approval Committee (CSAC) Vote: July 2014
8. Board of Directors Vote: August 2014
9. Appeals: September 2014

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

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

1. Importance to Measure and Report

(1a. High Priority, 1b. Opportunity for Improvement, 1c. Measure Intent)

IM.1. High Priority: **H-20; M-1; L-0; I-0** IM.2. Opportunity for Improvement: **H-10; M-10; L-0; I-1** IM.3. Measure Intent: **H-16; M-5; L-0; I-0** Overall Importance: **H-16; M-5; L-0; I-0**

Rationale:

- The Committee agreed that Acute Myocardial Infarction (AMI) is a high-priority area for measurement because it is a common condition that drives spending in hospitals.
- The Committee questioned the opportunity for improvement because the inner quartile of performance gets very narrow after risk adjustment. The developers responded that this measure is intended to be paired with quality measures and that the opportunity for improvement must be considered with the opportunity to improve the quality of care when factoring in the cost of the care provided.
- Additionally, the Committee was concerned with the attribution of post-acute expenses to the admitting hospital. The developers responded that it is critical to capture those costs because the current system is setup to incentivize pushing those payments out into the post-discharge time period. Hospitals can act as catalysts in their communities for improving care and health decision-making.
- The Committee raised a question about the episode definition as 30 days from the date of admission and the potential need for alignment with the Medicare Spending per Beneficiary (MSPB) measure that defines a period of 30 days post-discharge. The developers responded that these specifications are aligned with a corresponding AMI mortality measure to be used together to assess value.

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

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

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.

3. Feasibility: H-18; M-3; L-0; I-0

(3a. Byproduct of Care Processes; and 3b. Electronic sources; and 3c. Data Collection Strategy)

Rationale:

- The Committee had no concerns about the feasibility of the measure.

4. Use and Usability: H-12; M-7; L-2; I-0

(4a. Accountability/transparency (used in accountability w/in 3 yr, public reporting w/in 6 yr, or if new - credible plan); and 4b. Improvement – progress demonstrated (if new - credible rationale); and 4c. Unintended Consequences - benefits outweigh evidence of unintended negative consequences (to patients/populations); and 4d. Measure Deconstruction – can be deconstructed to facilitate transparency and understanding)

Rationale:

- The Committee raised concern about the number of hospitals falling in the "average" range for the measure – 78 percent. 15 percent were rated "high" and 7 percent "low".
- The Committee did appreciate the data breakdown provided to hospitals as a result of the measure.

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: Y-12; N-9

6. Public and Member Comment: April 21, 2014 – May 21, 2014

- Several supportive comments for the measure were received, with commenters indicating that the measure addresses an area of high morbidity, mortality, and healthcare costs. Commenters stated that information shared by CMS with hospitals will allow for identification of high/low cost areas and focused improvement. Additionally, commenters raised several issues with the measure, which were discussed during the in-person meeting:
    - Appropriateness of attribution approach
        - Commenters stated that attributing the cost of the entire episode to the admitting hospital may be inappropriate to attribute the cost of the episode to the hospital as much of the care happens in an outpatient setting. Commenters stated that measures should assess processes and outcomes over which the measured entity (e.g., hospital, physician group) can exercise a reasonable level of control, and that these measures may be more appropriate for an organization accepting bundled payments on behalf of all measured entities.
        - The Committee acknowledged this concern; however, the Committee stated that

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

increasingly hospitals are responsible for care delivered up to 30 days after discharge. Consequently, hospitals are in the unique position of being able to push coordination of care, and this measure may serve as an impetus for this to occur.

- Adequacy of risk adjustment model
  - Several commenters stated that the low r-squared values for the measure (0.05) indicated that the risk model did not account for enough of the variation in measure scores and may not adequately account for patient case mix and severity. Moreover, commenters believe that the low level of reliability demonstrated illustrated another fundamental flaw of both measures—that they fail to adequately account for complicating conditions that patients have prior to an episode of care.
  - The developers explained 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.
- Approach to addressing transfer patients
  - Several commenters stated concern that the initial admitting hospital would be attributed cost for the episode when transferring patients to a second hospital, as the initial admitting hospital may have little control over the care that happens after the transfer.
  - The Committee acknowledged this concern; however, the Committee stated that increasingly hospitals are responsible for care delivered up to 30 days after discharge.
- Risk adjustment for socio-demographic factors
  - Several commenters stated that the risk adjustment models for the measures should capture socio-demographic factors, as there is robust evidence that such factors affect health outcomes, including resource use.
  - NQF acknowledged these concerns and clarified that NQF is in the early stages of reviewing our policy on risk adjusting for socio-demographic factors. The report referenced is a draft report that has recently been reviewed during an NQF member and public comment period; the recommendations have not yet been finalized. As such, we ask that Committees continue to evaluate measures according to our current guidelines, that measures not be adjusted for socio-demographic variables. If in the future the recommendations for adjusting for socio-demographic variables become NQF policy, measures needing this adjustment will be updated and reviewed by the Committee through measure maintenance.
  - The Committee acknowledged that the timing of the NQF risk adjustment report is not ideal; however, given the current NQF policy on adjusting for sociodemographic variables, the Committee requested that a recommendation be issued with the measure that when reported, the results should be stratified by sociodemographic variables.

- After considering all comments and thorough discussion, the Committee requested the opportunity to revote on endorsement for the measure. The results of that vote are below:
  - Yes- 14; No-7
- The measure is recommended for endorsement and pursuant with NQF process will be posted for NQF

#2431 Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for Acute Myocardial Infarction (AMI)
member voting.
7. Consensus Standards Approval Committee (CSAC) Vote: July 2014
8. Board of Directors Vote: August 2014
9. Appeals: September 2014

#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] **1. Importance to Measure and Report** *(IM.1. High Priority; IM.2. Opportunity for Improvement; and IM.3. Measure Intent)* IM.1. High Priority: **H-14; M-4; L-3; I-0** IM.2. Opportunity for Improvement: **H-11; M-9; L-1; I-0** IM.3. Measure Intent: **H-11; M-9; L-1; I-0** Overall Importance: **H-8; M-13; L-0; I-0** Rationale: • The Committee agreed that Heart Failure (HF) is a high-priority area for measurement because it is a common condition that drives spending in hospitals and systems.
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

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

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.

3. Feasibility: H-16; M-3; L-0; I-0

(3a. Byproduct of Care Processes; and 3b. Electronic sources; and 3c. Data Collection Strategy)

Rationale:

- The Committee had no concerns about the feasibility of the measure.

4. Use and Usability: H-4; M-10; L-6; I-1

(4a. Accountability/transparency (used in accountability w/in 3 yr, public reporting w/in 6 yr, or if new - credible plan); and 4b. Improvement – progress demonstrated (if new - credible rationale); and 4c. Unintended Consequences - benefits outweigh evidence of unintended negative consequences (to patients/populations); and 4d. Measure Deconstruction – can be deconstructed to facilitate transparency and understanding)

Rationale:

- The Committee had no concerns about the Use and Usability of the measure.

5. Related and Competing Measures

- No related or competing measures noted.

Standing Committee Recommendation for Endorsement: Y-10; N-11

6. Public and Member Comment: April 21, 2014 – May 21, 2014

- Several supportive comments for the measure were received, with commenters indicating that the measure addresses an area of high morbidity, mortality, and healthcare costs. Commenters stated that information shared by CMS with hospitals will allow for identification of high/low cost areas and focused improvement. Additionally, commenters raised several issues with the measure, which were discussed during the in-person meeting:
    - Appropriateness of attribution approach
        - Commenters stated that attributing the cost of the entire episode to the admitting hospital may be inappropriate to attribute the cost of the episode to the hospital as much of the care happens in an outpatient setting. Commenters stated that measures should assess processes and outcomes over which the measured entity (e.g., hospital, physician group) can exercise a reasonable level of control, and that these measures may be more appropriate for an organization accepting bundled payments on behalf of

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

all measured entities.

- The Committee acknowledged this concern; however, the Committee stated that increasingly hospitals are responsible for care delivered up to 30 days after discharge. Consequently, hospitals are in the unique position of being able to push coordination of care, and this measure may serve as an impetus for this to occur.

- Adequacy of risk adjustment model
  - Several commenters stated that the low r-squared values for the measure (0.03) indicated that the risk model did not account for enough of the variation in measure scores and may not adequately account for patient case mix and severity. Moreover, commenters believe that the low level of reliability demonstrated illustrated another fundamental flaw of both measures—that they fail to adequately account for complicating conditions that patients have prior to an episode of care.
  - The developers explained 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.
- Approach to addressing transfer patients
  - Several commenters stated concern that the initial admitting hospital would be attributed cost for the episode when transferring patients to a second hospital, as the initial admitting hospital may have little control over the care that happens after the transfer.
  - The Committee acknowledged this concern; however, the Committee stated that increasingly hospitals are responsible for care delivered up to 30 days after discharge.
- Risk adjustment for socio-demographic factors
  - Several commenters stated that the risk adjustment models for the measures should capture socio-demographic factors, as there is robust evidence that such factors affect health outcomes, including resource use.
  - NQF acknowledged these concerns and clarified that NQF is in the early stages of reviewing our policy on risk adjusting for socio-demographic factors. The report referenced is a draft report that has recently been reviewed during an NQF member and public comment period; the recommendations have not yet been finalized. As such, we ask that Committees continue to evaluate measures according to our current guidelines, that measures not be adjusted for socio-demographic variables. If in the future the recommendations for adjusting for socio-demographic variables become NQF policy, measures needing this adjustment will be updated and reviewed by the Committee through measure maintenance.
  - The Committee acknowledged that the timing of the NQF risk adjustment report is not ideal; however, given the current NQF policy on adjusting for sociodemographic variables, the Committee requested that a recommendation be issued with the measure that when reported, the results should be stratified by sociodemographic variables.

- After considering all comments and thorough discussion, the Committee requested the opportunity to revote on endorsement for the measure. The results of that vote are below:
  - Yes- 13; No-8

#2436 Hospital-level, risk-standardized payment associated with a 30-day episode-of-care for heart failure (HF)
• The measure is recommended for endorsement and pursuant with NQF process will be posted for NQF member voting.
7. Consensus Standards Approval Committee (CSAC) Vote: July 2014
8. Board of Directors Vote: August 2014
9. Appeals: September 2014