TO: Consensus Standards Approval Committee (CSAC)

FR: Taroon Amin, Ashlie Wilbon, Lindsey Tighe, and Evan Williamson

RE: Cost and Resource Use Member Voting Results

DA: October 2, 2013

The CSAC will review recommendations from the Cost and Resource Use project at its October 8th conference call.

This memo includes a summary of the project, recommended measures, and themes identified from and responses to the public and member comments.

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 September 23.

Accompanying this memo are the following documents:

1. **Cost and Resource Use Draft Report** The draft report has been updated to reflect the changes made following Steering 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 within the comments received. This table lists 59 comments received and the NQF/Steering Committee responses.

CSAC ACTION REQUIRED
The CSAC will be asked to consider next steps, including CSAC voting on one candidate consensus standard.

Cost and Resource Use Measures Recommended for Endorsement:

- 2158 -- Payment-Standardized Medicare Spending Per Beneficiary (MSPB)
    - Measure put forward for member vote with split results.

Cost and Resource Use Measures Not Recommended:

- 2165 -- Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries
    - Due to lack of approval at steering committee, measure not put forward for member vote.

BACKGROUND

To expand the portfolio of endorsed cost and resource use measures that could be used as building blocks toward understanding efficiency and value, NQF conducted foundational work on cost and resource use definitions and the NQF Resource Use Measure Evaluation Criteria over the last several years. This prior work was captured in the final and technical reports and yielded the first eight endorsed cost and resource use measures in the NQF portfolio. The first consensus development project on cost and resource use measures served as the foundation for this project, which entails evaluation of non-condition specific measures of total cost, using per-capita or per-hospitalization approaches

DRAFT REPORT

The Cost and Resource Use Draft Report presents the results of the evaluation of two measures. One measure was recommended for endorsement by the steering committee as a voluntary consensus standard suitable for accountability and quality improvement and one measure was not recommended on a close steering committee vote. The measures were evaluated against the 2011 version of the measure evaluation criteria.

	MAINTENANCE	NEW	TOTAL
Measures considered	0	2	2
Withdrawn from consideration	0	0	0
Recommended	0	1	1
Not recommended	0	1	1
Reasons not Recommended	N/A	Overall- 1	

COMMENTS AND THEIR DISPOSITION

NQF received 59 comments from 25 organizations (including 20 member organizations) and individuals pertaining to the general draft report and to the measures under consideration.

A table of comments submitted during the comment period, with the responses to each comment and the actions taken by the Steering Committee and measure developers, is posted to the Cost and Resource Use project page under the Public and Member Comment section.

Comment Themes and Committee Responses

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

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

Comments on the General Draft Report

NQF received a number of comments on the general draft report, many of which focused on the importance of cost and resource use measures, and in particular, the need for pairing cost and resource

use measures with quality measures in order to have measures of efficiency and/or value. The commenters stated the importance of efficiency and/or value measures for assisting consumers in determining where to access the highest value healthcare.

Other comments included:

- Agreement that multiple risk adjustment methodologies within a single resource use measure must have data demonstrating comparable measure results
- Insistence that methodologies underlying all cost and resource use measures be soundly demonstrated

Steering Committee Response:
The Resource Use Steering Committee agrees with the commenters that measures of efficiency, and ultimately value are critical tools to assist consumers in determining where to access the highest value care. Further, members of the Committee also agreed that measures of resource use, and ultimately efficiency are needed to shift the US health care system away from fee-for-service and toward a value-based system. These measures submitted by CMS are an important first step toward this goal since measures of total cost should demonstrate that they are important to measure, have scientifically acceptable properties, and are usable and feasible. The Steering Committee encourages CMS to continue to refine the methodologies used to be more valid indicators of resource use and to explore the mechanism in which these resource use measures can be linked with quality performance to ensure that patients are selecting the highest value providers, not simply the least costly. The Steering Committee specifically recommends that providers' resource use performance should be compared to only those providers with equal or higher quality performance. The Committee urged that outcome and patient experience measures should be used when comparing resource use performance when available.

Measure Specific Comments: Measure Not Recommended

2165: Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries

This measure received comments from eighteen organizations/ individuals. Several commenters shared support for the concept and intent of the measure, urging CMS to make revisions to the attribution approach, risk adjustment algorithm, reliability and validity of the measure and to bring the measure back to NQF for endorsement, as this is an area where measures are needed and would provide insight into the costs of healthcare to Medicare. In addition to the support for the concept and intent of the measure, one commenter also acknowledged "provider concerns over the attribution of total cost of care to primary care physicians who are not part of an organized health system. But purchasers have come to expect care to be coordinated among providers and see the need to incentivize such coordination. Moreover, measures such as this one will help primary care physicians to understand the cost implications of their referral recommendations."

Steering Committee Response:
The Steering Committee raised several concerns with the construct of the measure, which the developer has been working to analyze and address during the past few months. Responses to the committee's concerns and additional analysis performed by the developer were shared with the committee on their August 28th call. The Steering Committee had the opportunity to review all comments and the developer's analysis and re-affirmed their decision to not recommend the measure for endorsement.

NQF will work with the developers to determine when it can next be reviewed for endorsement. This would happen when the next Cost and Resource Use project is scheduled.

The Steering Committee unanimously agreed that cost and resource use measures must be paired with quality measures in order to understand and make decisions about care. The Committee agrees with the commenters that measures of efficiency, and ultimately value are critical tools needed to improve the efficiency of US health care system, specifically encouraging shared accountability and team-based care.

The Steering Committee acknowledged the consumer perspective that care should be coordinated among providers; however, the Steering Committee was split over the idea that it may be inappropriate to hold primary care providers accountable for the cost of care provided to patients by other specialists, through inpatient care or through post-acute care, as primary care providers have limited ability to control these costs. In the current state of care delivery, health care is accessed in many ways. Many patients select their own primary and specialty care physicians, making decisions to see providers on their own, without coordination with their PCP or PCP group. Several Committee members stated that this may be appropriate in markets with integrated care delivery networks or where patients identify with a PCP or PCP group voluntarily or by assignment; however, in the current fragmented state of care delivery this attribution approach is not preferred Several other committee members stated that this level of accountability for providers is the entire rationale for the measure and should help push providers to be better organized to reduce costs.

The remaining comments addressed several themes listed below, along with a description of the comments received.

Attribution

- Several commenters agreed with the Committee that primary care physicians or specialists who may be attributed patients because they provided primary care services to that patient have limited ability to control the cost of care provided to patients by other specialists, through inpatient care or through post-acute care. The majority of commenters agreed that it may be inappropriate to hold these providers accountable for these costs of care. The commenters also agreed that this may be appropriate in markets with integrated care delivery networks; however, in the current fragmented state of care delivery this attribution approach is not supported.
- Additionally, several commenters shared the Steering Committee concerns that patients and their associated costs may potentially be attributed to specialists who provide primary care services that are Medicare allowable charges and questioned the appropriateness of this.
- Several commenters shared the Steering Committee concern that visits with non-physician providers (PAs and NPs) are not taken into account in the attribution model until the second stage, as non-physician providers are increasingly delivering more primary care.
- Given the various concerns about the attribution approach, several commenters called into question the reliability and validity of the measure, noting the Steering Committee's split vote as to whether the measure was in fact valid.

Steering Committee Response:

The Steering Committee acknowledged many of the same concerns with the attribution approach. The Steering Committee stated concern that patients and their associated costs may potentially be attributed to specialists who provide primary care services that are Medicare allowable charges. This is particularly significant in the case of patients who receive long-term care for chronic conditions, who may receive many primary care services from specialists treating them for their chronic conditions, who are then attributed to a medical group practice based on the plurality of Medicare allowable charges. The Committee noted the distinction that specialists can provide primary care services through visits other than primary care visits.

The Committee was ultimately split on the concern that physicians have little ability to control the cost of care provided to patients by other specialists, through either inpatient care or post-acute care. Several Steering Committee members raised concern that it may be inappropriate to hold these providers accountable for these costs of care. Further, several Committee members stated that this may be appropriate in markets with integrated care delivery networks; however, in the current fragmented state of care delivery this attribution approach is not preferred. On the other hand, several other committee members stated that this level of accountability for providers is the entire rationale for the measure and should help push providers to be better organized to reduce costs.

The Steering Committee agreed with commenters that there are issues with both the first and second stage of the attribution approach. In the first stage, visits with non-physician providers (PAs and NPs) are not taken into account in the attribution model until the second stage, as non-physician providers are increasingly delivering more primary care. The Committee strongly encourages CMS to include non-physician providers in the first stage of the attribution approach. Further, primary care services as defined by this measure may not always represent actual primary care visits by primary care providers. The Committee encourages CMS to update this attribution approach.

Exclusions

- One commenter expressed concern that the exclusions of death and Medicare Advantage beneficiaries impact the usability of the measure.
- One commenter expressed concern that Medicare Part D (prescription medications) was excluded from the measure.

Steering Committee Response:
The committee was split on the reliability and validity of this measure but ultimately agreed that a number of issues, including the exclusions of death needed to be addressed before recommending this measure for endorsement. Additionally, Medicare Part D payment is an important area for measurement and improvement. CMS should consider approaches to including this data for beneficiaries with Part D coverage.

Reliability

- One commenter requested that the measure developer not publically report results for any provider group with reliability scores less than 0.70.
- One commenter stated that the measure is only reliable for groups of 25 or more eligible professionals; however, nearly half of all Medicare physicians practice in groups of fewer than 10 eligible professionals. As the measure will be used as part of CMS' value-based modifier calculation, the commenter questioned how this will impact smaller physician groups and solo practitioners.

Steering Committee Response:
While NQF does not require a specific cut-off for reliability testing, the Committee does encourage CMS to report information on provider groups that have adequate reliability in performance score and sample size. This measure should only be used for 25 or more eligible professions since this is the scope of measure testing.

Risk Adjustment

- One commenter expressed concern that the risk adjustment model might not adequately capture the differences in patient population for different specialties, particularly those who treat patients with uncommon and very severe diseases.
    - Comment #3253: We appreciate the opportunity to comment on the measure #2165 (Payment-Standardized Total Per Capita cost Measure for Medicare Fee-for-Service Beneficiaries). We have some concerns about this measure and its potential for use as a component of the value-based modifier. From the measure description

and information provided, it is unclear how this measure would be applied. We are concerned about the broad nature of this measure and the fact that it looks across different specialties rather than within each specialty. We understand that the risk adjustment takes into account complexity of disease; however, we are concerned that the risk adjustment model might not adequately capture the differences in the patient population for different specialties. We are concerned that certain specialties, particularly cognitive specialists like rheumatology caring for patients with uncommon and very severe diseases, as a whole might fare worse than others if this measure is applied across specialties. In addition, we reviewed the risk adjustment model and do not believe it adequately captures the scope and complexity of conditions that rheumatologists care for. The exclusion of consideration of specific patient populations in the risk adjustment model would put providers or centers who treat a large number of these patients at a disadvantage. We would urge any assessment of providers for efficiency to look within a specialty rather than across specialties and that the risk adjustment model be thoroughly reviewed through specialty societies.

- Several commenters stated that the HCC model, which was developed for the Medicare Advantage program, does not adequately account for risk for purposes of analyzing physician group resource use, as it was designed to risk adjust large patient populations for insurance rate determination.

Steering Committee Response:
The Committee generally agreed that while this HCC-risk adjustment model was developed for Medicare Advantage it was appropriate but weak in this application. The HCC model does not include as many diagnostic categories as many commercially available risk adjustment models and therefore may not be as accurate in assigning the appropriate risk categories for rare conditions. However, given the broad use of HCCs across Medicare programs the Committee agreed that this approach was sufficient for this application.

ACTION TAKEN:

- After review and discussion of the comments on this measure, the Committee narrowly re-affirmed their decision to not recommend this measure for endorsement (12 recommend, 13 not recommend).
- Due to lack of approval at steering committee, the measure was not put forward for member vote.

Measure Specific Comments and Votes: Recommended measure

2158: Payment-Standardized Medicare Spending per Beneficiary (MSPB)

This measure received comments from twenty organizations/ individuals. Three of these comments were supportive, noting that this measure is an important first step "towards an optimal measure of hospital resource use." One commenter noted that the measure "does have methodologic concerns but its intent is clear and necessary."

The remaining comments addressed several themes listed below, along with a description of the comments received.

Exclusion of Deaths

- Two commenters questioned the exclusion of deaths, noting that they "believe the measure would be a stronger measure of costs if patient deaths were included."
- One commenter supported the exclusion of deaths and also called for the exclusion of hospice payments in order to "maintain the internal consistency of the measure."

Steering Committee Response:
The Resource Use Steering Committee generally agrees with the Commenter that the inclusion of episodes where the patient dies would create a stronger measure. End of life care is a high-cost area for the Medicare program and is important for measurement and improvement. The developer acknowledges that the exclusion was finalized through notice and comment rulemaking based on the fact that these are incomplete episodes where significant data could be missing, but that CMS will consider including episodes in which the beneficiary dies in future updates to the MSPB measure.

Exclusions

- One commenter expressed concern that exclusion of transfer patients from other acute care facilities may affect a larger portion of PPS-exempt Cancer Center patient admissions when compared to PPS Hospital admissions.
- One commenter expressed concern that exclusion of transfer patients could remove more seriously ill patients, which represent significant opportunities for reduced spending.
- One commenter stated that inclusion of Medicare Part D data would result in a stronger measure.

Steering Committee Response:
The committee agrees that additional analysis would need to be conducted to determine the transferability of validity results to a cancer patient population. As specified, the measure currently excludes cancer hospitals.

Furthermore, additional analysis on risk adjustment approach specific to PPS-Exempt Cancer Centers' patient population and the 90-day look back period would need to be conducted before the measure was specified for a cancer patient population.

The Resource Use Steering Committee generally agrees with the commenter that facilities being held responsible for the utilization and associated costs for patients that they transfer to other facilities would foster better collaboration resulting in more efficient and effective care. This collaboration fits with the philosophy of holding a facility responsible for care delivered up to 30 days post discharge.

The Committee agrees that inclusion of Part D data would create a stronger measure. They recognized the limitation in the availability of these data; however, they encouraged the measure developers to consider additional strategies to include these costs in the future.

Attribution

- Commenters cautioned that this measure is only suitable for reporting at the facility level and should not be analyzed or reported at the individual clinician level. Commenters stated concern that the measure has not been tested or specified for this analysis at the individual clinician level.
- Commenters also agreed with the Steering Committee recommendation that this measure be reported with quality measures, in order to provide meaningful information about the efficiency of health care delivery.

Steering Committee Response:
The Steering Committee unanimously agreed that cost and resource use measures must be paired with quality measures in order to understand and make decisions about care. The Committee agrees with the commenter that measures of efficiency, and ultimately value are critical tools needed to improve the efficiency of US health care system, specifically encouraging shared accountability and team-based care.

Measures endorsed by NQF are only endorsed for use at the specified level of analysis that the measure developer has provided testing for; in this case, that would be at the facility level. The Steering Committee has only recommended this measure for endorsement for analysis at the facility level.

Risk Adjustment

- Several commenters stated concern that the risk adjustment methodology was not valid for the following reasons:
    - Lack of a socio-economic status (SES) adjustment (i.e. dual-eligible status).
        - Comment #3249: CMS's analysis demonstrates that dual-eligible patients have $859 more spending per episode than other patients. The agency finds that including patient dual-eligible status as a risk adjuster marginally improved the fit (R-squared value) of the risk adjustment model. But, the same analysis also demonstrates that about 12% of hospitals would have their MSPB measure values change by more than 1 percentage point if dual-eligible status were included in the risk adjustment model. About 10.8% of hospital scores would decrease by between 1 and 3 percentage points. Nevertheless, CMS chose to not include a dual-eligible adjustment in the measure.
    - Testing results demonstrating clustering of large, urban, teaching hospitals that treat a large proportion of low income patients with higher MSPB index rates than their community hospital counterparts, possibly due to the risk adjustment not accounting for the ranges of

patient complexity that exist between and within MS-DRGs or that case mix is driving he differences in measure score.

- Comment #3258: The actual results of the MSPB suggest that the case mix adjustment isn't working properly. In Minnesota, for example, hospitals in urban areas have similar scores, clustering around .93. In Greater Minnesota, however, the scores are almost all less than .88. Because there are large differences in the types of conditions treated by urban and rural hospitals, it raises a concern that the case mix is driving the differences vs. actual differences in adjusted resource use.
- Comment #3261: We thank NQF for the opportunity to comment on the cost and resource use measure, #2158 Medicare Spending per Beneficiary (MSPB). We have several areas of concern with this measure that should be addressed prior to endorsement. In the MSPB results that have been published by the measure developer, there is a noticeable clustering of large, urban, teaching hospitals that treat a large proportion of low income patients with higher MSPB index rates than their community hospital counterparts. We believe this is due to insufficient severity adjustment in the measure that does not account for the ranges of patient complexity that exist between and within MS-DRGs. Since large, urban, teaching hospitals, with a large share of low income patients have the capability to treat more complex patients, whereas community hospitals often do not, they have a higher proportion of complex cases that require more hospital resources and are also more likely to have home care or skilled nursing care following the inpatient admission. This mix of more complex patients could be a contributing factor to the clustering being seen in the results. We would expect a more normal distribution of MSPB results across all hospitals if the measure were appropriately severity/risk adjusted and adjusted for outliers.

- Risk stratification using MDC criteria alone is inadequate and will introduce significant variability in the MSPB rating based upon patient-specific and diagnosis-specific factors that are not adequately encompassed in the MDC classification.
  - Comment #3271: NASS is concerned that the risk stratification for MSPB does not have adequate granularity to differentiate significant cost drivers. Specifically, the proposal to stratify cases by major diagnostic category (MDC).There is significant evidence that MDC classification does not accurately encompass the factors that contribute to cost of care, and there are significant inaccuracies in the administrative data that contributes to the MDC Classification. Risk stratification using MDC criteria alone is inadequate and will introduce significant variability in the MSPB rating based upon patient-specific and diagnosis-specific factors that are not adequately encompassed in the MDC classification. A more comprehensive classification that includes an algorithm that includes CPT codes and procedure specific information would be more useful than a stratification based upon MDC alone.
- Concern that the 90-day look-back period to capture a patient's comorbidities in order to determine the HCC score is insufficient.

Steering Committee Response:

The Committee reviewed the concern around the clustering of large, urban, teaching hospitals that treat a large proportion of low-income patients raised by the commenters. They voiced concerns on both sides of the issue of including SES adjustment with some members agreeing with the Commenter that disadvantaged patients with multiple complex conditions will require more resources to treat, while other members argued that including SES variables in the risk adjustment model will mask disparities in cost performance among different groups of patients. The Committee acknowledged that hospitals are legitimately held accountable for taking appropriate care of patients within the case mixes and making sure that these patients receive appropriate post-acute care, however, the availability of these support services will vary from community to community.

The Committee recommended that additional work be considered in this area, specifically the appropriateness of including dual-eligibility in risk adjustment models for resource use measurement.

The Committee considered the major diagnostic category (MDC) risk stratification criteria, specifically applying the risk adjustment within MDC to be generally appropriate for this application. For the purposes of performance measurement, factors that are included in the risk adjustment model should be present at the start of care – thus including procedure codes that occur during the measurement period would not be appropriate.

The Committee also expressed concern over the 90-day look back period but ultimately agreed that the performance of the models did have a slightly improved model fit over the models with a year of look-back.

ACTION TAKEN:

- After review and discussion of the comments on this measure, the Committee re-affirmed their decision to recommend this measure for endorsement (17-8).

NQF MEMBER VOTING RESULTS

The recommended measure was not approved with 43% of councils approving. Representatives of 42 member organizations voted; no votes were received from the Public/Community Health Agency Council. Results for each measure are provided below.

NQF Member Council	Voting Organizations	Eligible to Vote	Rate
Consumer	4	28	14%
Health Plan	4	15	27%
Health Professional	16	87	18%
Provider Organizations	7	134	5%
Public/Community Health Agency	0	33	0%
Purchaser	5	24	21%
QMRI	4	69	6%
Supplier/Industry	3	29	10%
All Councils	**43**	**419**	**13%**

2158: Payment-Standardized Medicare Spending per Beneficiary

Member Council	Yes	No	Abstain	Total Votes	% Approval*
Consumer	4	0	0	4	100%
Health Plan	4	0	0	4	100%
Health Professional	2	13	1	15	13%
Provider Organizations	2	5	0	7	29%
Public/Community Health Agency	0	0	0	0	
Purchaser	5	0	0	5	100%
QMRI	2	2	0	4	50%
Supplier/Industry	0	3	0	3	0%
All Councils	**19**	**22**	**1**	**42**	**45%**
Percentage of councils approving (>50%)					**43%**
Average council percentage approval					56%

*equation: Yes/ (Total - Abstain)

Voting Comments:

- AmeriHealth Caritas: Considering the number of issues raised during comment period, this measure as it stands should be considered a good starting off point for more accurate successor measures.
- Next Wave: We voted to approve the measure as specified and acknowledge that risk adjusting for dual eligibility status could mask performance opportunities. However, we also recognize that costs are consistently and measurably higher when treating patients with complex social disadvantages (language, literacy, low income, lack of informal supports, etc.) to achieve the same health outcomes. Additional costs include providing competent care across many different cultures, patient education that addresses basic literacy as well as health literacy, post-discharge transportation to fill prescriptions and get to MD visits, home visits to assess informal or provide direct post-discharge support in the Community, etc. When applying this measure for Value Based Purchasing, it should be stratified to compare hospitals with similar social determinate/population complexity. Otherwise, hospitals will be perversely rewarded for not admitting these high need patients and penalized if they do admit these high expected cost patients, likely increasing health disparities. This measure application issue should be placed on the agenda of the MAP.
- American College of Cardiology/ American Heart Association Performance Measure TF: This resource measure should be used in conjunction with a quality measure.
- American College of Cardiology: This measure should only be used in public reporting or provider incentive programs when tightly associated with related quality of care measures. No provider should ever be rewarded for lowering costs at the expense of patient care.
- American College of Surgeons: The American College of Surgeons (ACS) has concerns regarding the validity of the MSPB measure. The risk adjustment methodology does not adequately account for socioeconomic status (SES) and therefore does not consider the ranges of patient

complexity and circumstances that are out of the provider□s control. As a result, the lack of risk adjustment for SES may unfairly impact providers or facilities that care for disadvantaged populations. Additionally, we realize that the MSPB measure is being considered for endorsement as a facility-level measure but we encourage NQF to recommend against the implementation of this measure as a physician-level measure unless otherwise tested and specified at the physician-level. The ACS opposes reliance on the implementation of measures in the public domain which were endorsed by NQF for one level of measurement yet implemented at a different level of measurement for which they were not validated.

- American Hospital Association: While we commend CMS for carefully thinking through many aspects of this measure in developing it to respond to a specific legislative mandate in the ACA, we remain concerned that the risk adjustment is insufficient, the testing results are problematic and, as noted by many commenters, there are significant issues with some of the inclusion/exclusion decisions that likely mean this measure" does not yield results that are suitable for comparison purposes. Before receiving the imprimatur of an NQF endorsement we believe more work is needed. We are particularly concerned that there is no adjustment for factors that are totally outside the control of the hospital and individual clinician groups. These are factors related to the resources available in the community served by a hospital or physician group. We talk about these as adjustments for "socio-economic factors" and in a way they are but they are really adjustments for the fact that inadequate primary care wellness care and so forth exist in many communities of lower income and as a result Medicare patients end up in the hospital with many concerns that need to be addressed. We need to attend to these issues before this and many other measures receive NQF endorsement."
- America's Health Insurance Plans: While this measure represents a good start for assessing hospital cost and resource use, we encourage further refinement by CMS. Future iterations of this measure should remove existing exclusions such as deaths and hospitals transfers. These exclusions remove more seriously ill patients, who represent significant opportunities for reduced spending such as spending associated with potential overuse of services provided at end of life.

  We also recommend exercising caution when implementing this measure and drawing conclusions from its use. For example, excluding transfers would solidify fragmentation of accountability and remove motivation for referring and receiving hospitals to collaborate on more efficient and effective care.
- AAMC: We are concerned with the lack of socioeconomic status adjustment in this measure.
- Infectious Diseases Society of America: The Medicare Spending Per Beneficiary (MSPB) measure is concerning since it evaluates costs related to the totality of services furnished to a patient surrounding an inpatient hospitalization. This includes all Medicare Part A and Part B payments during the episode, which spans from 3 days prior to an index admission through 30 days post discharge, with certain exclusions. In the Voting Draft Report, it states, “The Steering Committee unanimously agreed that cost and resource use measures must be paired with quality measures in order to understand and make decisions about care.” Similarly, IDSA believes it is critical that cost measures have a more direct link to the quality measures used to assess value. Conclusions about the value of medical care will have little significance if the cost and quality measures on which they are based focus on different elements of care. The IDSA believes it is important for each physician caring for a patient to understand how he/she contributes to the patient’s total cost of care. However, it is not necessarily appropriate to hold each of these physicians accountable for the patient’s total cost of care. We fear not only

physicians being held responsible for decisions outside of their control, but unintended consequences such as physicians not ordering labs prior to prescribing antibiotics in order to minimize costs, which could lead to inappropriate antibiotic use, increased drug resistance, and harm to the patient, as well as the general public. Therefore, until this measure is fully developed and tested to accurately reflect care over which the individual clinician has control, IDSA cannot support endorsement of this measure.

APPENDIX: Measure Evaluation Summary Table

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

Recommended Measure

#2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB)

Steward: Centers for Medicare and Medicaid Services

Description: The MSPB Measure assesses the cost of services performed by hospitals and other healthcare providers during an MSPB hospitalization episode, which comprises the period immediately prior to, during, and following a patient's hospital stay. Beneficiary populations eligible for the MSPB calculation include Medicare beneficiaries enrolled in Medicare Parts A and B who were discharged from short-term acute hospitals during the period of performance.

Resource Use Measure Type: Per episode

Data Source: Administrative claims

Level of Analysis: Facility

Costing Method: Standardized pricing

Target Population: Senior Care

Resource Use Service Categories: Inpatient services: Inpatient facility services; Inpatient services: Evaluation and management; Inpatient services: Procedures and surgeries; Inpatient services: Imaging and diagnostic; Inpatient services: Lab services; Inpatient services: Admissions/discharges; Ambulatory services: Outpatient facility services; Ambulatory services: Emergency Department; Ambulatory services: Evaluation and management; Ambulatory services: Procedures and surgeries; Ambulatory services: Imaging and diagnostic; Ambulatory services: Lab services; Durable Medical Equipment (DME)

STEERING COMMITTEE MEETING [May 8-9, 2013]

Importance to Measure and Report: The measure meets the Importance criteria

(1a. High Impact: 1b. Performance Gap, 1c. Measure Intent)

1a. Impact: **H-23**; **M-2**; **L-0**; **I-0** 1b. Performance Gap: **H-12**; **M-12**; **L-1**; **I-0** 1c. Measure Intent: **Y-6**; **N-16**; **I-3**; **L-0**

1. Overall: **H-9**; **M-15**; **L-1**; **I-0**

Rationale: While evaluating the measure's importance to measure and report, the Committee agreed that the subcriteria were met and provided the following rationale:

#2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB)

- General agreement that healthcare cost is a high impact area of healthcare.
- Affordability of healthcare has been identified as an area of focus as part of the Triple Aim and under the National Quality Strategy.
- Inpatient costs are a major driver of total costs; capturing this may incentivize hospitals to examine causes of these expenditures.
    - Readmissions and Skilled Nursing Facility costs will be significant drivers of cost captured through this measure; these are high impact areas where Medicare spends the most money with respect to hospitalizations.
- Though the developers stated that a benefit of the measure would be to improve care coordination, the Steering Committee did not agree that the evidence submitted substantiated this claim.
- Though the measure was described as a cost measure, the Steering Committee clarified that this is a Medicare expenditure measure, which can be used as a proxy for cost.

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-14**; **L-1**; **I-0** 2b. Validity: **H-0**; **M-13**; **L-11**; **I-1**

Rationale: While evaluating the measure's scientific acceptability, the Committee agreed that the subcriteria were met and identified 5 major issues:

1) Reliability concerns relating to 30% of hospitals moving quintiles as demonstrated in the test, re-test results
2) Validity concerns relating to the exclusion of deaths and transfers
3) Concern regarding the construct validity testing results which demonstrated low correlation with measures of readmissions in heart attack, heart failure, and pneumonia
4) Concern regarding the look back period for the HCC risk adjustment model
5) Concern regarding the appropriateness of not incorporating the dual eligible population into the risk adjustment model

1) Reliability concerns relating to 30% of hospitals moving quintiles as demonstrated in the test, re-test results
    - Many Committee members expressed concern that the test, re-test results demonstrated that approximately 30% of hospitals in the lowest-spending quintile in one sample were not in the lowest-spending quintile in the next sample; similarly, approximately 30% of hospitals in the highest-spending quintile in one sample were not in the highest-spending quintile in the next sample.
    - Committee members questioned whether this level of reliability would be sufficient, particularly with respect to establishment of cutoff thresholds when the measure is reported.
    - The developer stated that Spearman rank correlation for a hospital across samples is 0.835, demonstrating a linear relationship between the rank of the hospitals in the test and re-test samples. This indicates that using a different random group of patients does not result in significant variation of the hospital's relative performance.

2) Validity concerns relating to the exclusion of deaths and transfers
    - Several Committee members expressed concern that the exclusion of deaths and transfer

patients from the measure is unnecessary.
    - Exclusion of deaths removes from the measure calculation some of the patients who use the highest resources and thus are the most expensive. Additionally, the Committee questioned the rationale for inclusion of hospice costs when deaths are excluded.
    - Exclusion of transfer patients accounts for approximately 5% of patients, and the rationale for excluding them is unclear. The Committee members stated that, given that the measure holds the hospital accountable for patients 30 days after discharge, it isn't clear why transfers are excluded. Additionally, the Committee members stated that exclusion of transfers may result in gaming of the measure, as hospitals may simply transfer high cost patients.
- The developer stated that deaths were excluded because of the bimodal distribution of costs, with an average cost for patients who die 40% higher than those patients who do not die. However, many episodes cost far under what was predicted, potentially because the patient died early in the episode and thus did not utilize resources.
- The developer stated that transfer patients were excluded because of difficulties with attributing the patients to a hospital.
- Several Committee members stated concern that the rationale provided by the developer for excluding deaths and transfers was insufficient and suggested the measure developer consider updating the measure to address this concern.

3) Concern regarding the construct validity testing results which demonstrated low correlation with measures of readmissions in heart attack, heart failure, and pneumonia
    - Several Committee members stated that high correlation between the MSPB measure and a measure capturing readmissions is expected because of the high cost of readmissions for these diseases.
    - The Committee stated concern that the testing results for the measure demonstrated weak correlations with the readmissions measures (0.08, 0.07, and 0.06 for heart attack, heart failure, and pneumonia readmission rates respectively), particularly because the developers used this to demonstrate validity of the measure.
    - The developer speculated that the weak correlation resulted from the fact that the MSPB measure assesses the cost to Medicare of all services performed by hospitals and other healthcare providers during an MSPB episode; as a result, a hospital's MSPB measure value is driven by both acute and post-acute spending.
    - Several Committee members stated that the rationale provided by the developer on why spending and readmissions should be correlated needs to be substantiated by further testing, as these results provided demonstrated weak validity of the measure.
    - The developer also submitted validity testing results for the 30-day MSPB post-discharge window, demonstrating a positive correlation (0.13) between MSPB measure values and the percent of beneficiaries with multiple episodes. This analysis was intended to demonstrate that the measure is sensitive to the length of the 30-day post discharge window. The analysis aimed to analyze whether hospitals whose beneficiaries incurred multiple 30-day episodes performed better on the measure by virtue of the beneficiaries' care having been split into more episodes that were less expensive individually. The analysis, however, found that high cost hospitals are more likely to have beneficiaries with multiple episodes. This indicates that the 30-day window is not strongly affecting the measure.
        - Additionally, the developer further explored the validity of the 30-day MSPB post-

discharge window by testing rank correlation against a 90-day window. The developer found a positive rank correlation (0.897), suggesting that hospitals with high MSPB measures using the 30-day window also had high MSPB measures using the 90-day window.

4) Concern regarding the look back period for the HCC risk adjustment model
   - Several Committee members stated the concern that the HCC risk adjustment model only captures health status variables derived from claims during the 90 days prior to the start of an episode. Committee members stated that accuracy of the HCC model drops off dramatically with less than 7 months of data; 12 months of data is the gold standard.
   - The developer stated that testing was done to evaluate the health status variables in the risk adjustment model by using one year of data prior to the start of an episode rather than 90 days. The developer found that 6% of episodes are dropped, and the R-squared value actually decreases from 0.4621 (90 days data) to 0.4601 (one year data). Summarized, the developer found that capturing 90 days of data rather than one year of data resulted in no significant trade-off between the number of episodes included and the model fit.

5) Concern regarding the appropriateness of not incorporating the dual eligible population into the risk adjustment model
   - Steering Committee members voiced opinions on both sides of this issue, with some stating that dual eligible patients should be included in the risk adjustment model and others stating that they should not.
     - Those in favor of including dual eligible patients in the model stated concern that a potentially significant unintended consequence of not including dual eligible patients in the risk adjustment model would be the refusal of hospitals to accept dual eligible patients, as they are known to be higher cost than traditional Medicare patients.
     - Those opposed to including dual eligible patients in the model stated concern that adjusting for dual eligible status would mask any disparities in the cost of care for these patients.
   - The developer stated that although dual eligible patients are included in the measure population, a dual eligible risk adjuster is not currently included in the risk adjustment model.
   - A commenter from the public stated that dual eligible patients share characteristics beyond socioeconomic status, such as multiple chronic conditions, complex societal issues, and disparities in healthcare literacy. They are a population with chronic, complex disease that needs to be accounted for, particularly as relates to the impact on Safety Net hospitals within the context of this measure.

Additional Issues:

- *MS-DRG Regression.* Using Medicare Severity - Diagnostic Related Groups (MS-DRG) variables in the regression has the potential to mask variation attributable to quality of care, as patients can be bumped into a higher DRG through comorbidities or complications. The Committee questioned whether the standardized core DRG had been subtracted before the regression to see how much variance in the rest of the payments are explained by the other health status variables included in the risk adjustment model. The developer is willing to do this analysis.
- *Fiscal year payment rates.* Measure uses payment rates at the time of the claim (for the relevant fiscal year); potential for bias exists if admission rates vary significantly between fiscal years

#2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB)
between hospitals. • *Pre- and post-hospitalization services.* Several Committee members stated the concern that the major sources of variation between hospitals after risk adjustment are the pre-hospitalization and the post-hospitalization care; the Committee questioned whether the measure allowed for understanding of which sets of post-acute services result in higher cost when the measure is calculated.
3. Feasibility: H-23; **M-1**; **L-0**; **I-0** *(3a. Byproduct of Care Processes; and 3b. Electronic sources; and 3c. Data Collection Strategy)* Rationale: While evaluating the measure's feasibility, the Committee agreed that the subcriteria were met and provided the following rationale: • Data for the measure is being collected and is available. • Data is generated electronically. • The Committee generally agreed that the measure is feasible to implement.
4. Usability: H-6; M-15; L-3; 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: While evaluating the measure's usability and use, the Committee agreed that the subcriteria were met and provided the following rationale: • The Committee largely agreed that the measure will be most useful when paired with quality outcome measures. • The measure is in use for accountability purposes. • The Committee expressed concern that many hospitals may not have the analytic capacity to understand the data and understand the impact of care outside of the hospitalization on the measure result.   ○ The Committee recommended that the reports from CMS provide hospitals with analysis to allow hospitals to identify cost drivers outside of the hospitalization.   ○ The Committee recommended that CMS provide the hospitals with information on which post-acute care providers are using the most resources, so that hospitals can partner with providers who are utilizing fewer resources and providing quality care. • The developer stated that hospitals are provided with several different files to understand costs, including hospital-specific reports on its performance on the MSPB measure and patient-level data. The reports also provide comparison of a hospital's performance compared to other hospitals in the same state or across the nation, and provide a breakdown of spending by claim type. • From a consumer's perspective, the small variation in performance will make it difficult for the consumer to distinguish the best performers. The data is presented in a way that may be challenging for a consumer to deconstruct. • The developer stated that downloadable files are available online which will provide more detail on the measure results for consumers.

#2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB)
Unintended Consequences: While evaluating the measure's usability, the Committee identified the following potential unintended consequences: • Consumers may choose the most expensive hospital, believing that increased cost corresponds to higher quality healthcare. • Hospitals may transfer patients based on expected high expenditures post-discharge, resulting in the patient being excluded from the measure.
5. Related and Competing Measures • No related or competing measures noted.
Steering Committee Recommendation for Endorsement: **Y-15**; **N-10** Rationale: • The measure focus is high impact; healthcare costs in the United States are very high. • The measure allows hospitals to begin looking at and understanding cost; when paired with quality outcome measures, this will help hospitals gain an understanding of the value and efficiency of healthcare services provided.
Public and Member Comment [July 9 – August 7, 2013] This measure received comments from twenty organizations/ individuals. Three of these comments were supportive, noting that this measure is an important first step "towards an optimal measure of hospital resource use." One commenter noted that the measure "does have methodologic concerns but its intent is clear and necessary." The remaining comments addressed several themes listed below, along with a description of the comments received. **Exclusion of Deaths** • Two commenters questioned the exclusion of deaths, noting that they "believe the measure would be a stronger measure of costs if patient deaths were included." • One commenter supported the exclusion of deaths and also called for the exclusion of hospice payments in order to "maintain the internal consistency of the measure." **Steering Committee Response:** The Resource Use Steering Committee generally agrees with the Commenter that the inclusion of episodes where the patient dies would create a stronger measure. End of life care is a high-cost area for the Medicare program and is important for measurement and improvement. The developer acknowledges that the exclusion was finalized through notice and comment rulemaking based on the fact that these are incomplete episodes where significant data could be missing, but that CMS will consider including episodes in which the beneficiary dies in future updates to the MSPB measure. **Exclusions** • One commenter expressed concern that exclusion of transfer patients from other acute care facilities may affect a larger portion of PPS-exempt Cancer Center patient admissions when compared to PPS Hospital admissions. • One commenter expressed concern that exclusion of transfer patients could remove more seriously ill patients, which represent significant opportunities for reduced spending.

#2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB)

- One commenter stated that inclusion of Medicare Part D data would result in a stronger measure.

Steering Committee Response:
The committee agrees that additional analysis would need to be conducted to determine the transferability of validity results to a cancer patient population. As specified, the measure currently excludes cancer hospitals.

Furthermore, additional analysis on risk adjustment approach specific to PPS-Exempt Cancer Centers' patient population and the 90-day look back period would need to be conducted before the measure was specified for a cancer patient population.

The Resource Use Steering Committee generally agrees with the commenter that facilities being held responsible for the utilization and associated costs for patients that they transfer to other facilities would foster better collaboration resulting in more efficient and effective care. This collaboration fits with the philosophy of holding a facility responsible for care delivered up to 30 days post discharge.

The Committee agrees that inclusion of Part D data would create a stronger measure. They recognized the limitation in the availability of these data; however, they encouraged the measure developers to consider additional strategies to include these costs in the future.

Attribution

- Commenters cautioned that this measure is only suitable for reporting at the facility level and should not be analyzed or reported at the individual clinician level. Commenters stated concern that the measure has not been tested or specified for this analysis at the individual clinician level.
- Commenters also agreed with the Steering Committee recommendation that this measure be reported with quality measures, in order to provide meaningful information about the efficiency of health care delivery.

Steering Committee Response:
The Steering Committee unanimously agreed that cost and resource use measures must be paired with quality measures in order to understand and make decisions about care. The Committee agrees with the commenter that measures of efficiency, and ultimately value are critical tools needed to improve the efficiency of US health care system, specifically encouraging shared accountability and team-based care.

Measures endorsed by NQF are only endorsed for use at the specified level of analysis that the measure developer has provided testing for; in this case, that would be at the facility level. The Steering Committee has only recommended this measure for endorsement for analysis at the facility level.

Risk Adjustment

- Several commenters stated concern that the risk adjustment methodology was not valid for the following reasons:
    - Lack of a socio-economic status (SES) adjustment (i.e. dual-eligible status).
        - Comment #3249: CMS's analysis demonstrates that dual-eligible patients have $859 more spending per episode than other patients. The agency finds that including patient dual-eligible status as a risk adjuster marginally

#2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB)

improved the fit (R-squared value) of the risk adjustment model. But, the same analysis also demonstrates that about 12% of hospitals would have their MSPB measure values change by more than 1 percentage point if dual-eligible status were included in the risk adjustment model. About 10.8% of hospital scores would decrease by between 1 and 3 percentage points. Nevertheless, CMS chose to not include a dual-eligible adjustment in the measure.

- Testing results demonstrating clustering of large, urban, teaching hospitals that treat a large proportion of low income patients with higher MSPB index rates than their community hospital counterparts, possibly due to the risk adjustment not accounting for the ranges of patient complexity that exist between and within MS-DRGs or that case mix is driving he differences in measure score.
  - Comment #3258: The actual results of the MSPB suggest that the case mix adjustment isn't working properly. In Minnesota, for example, hospitals in urban areas have similar scores, clustering around .93. In Greater Minnesota, however, the scores are almost all less than .88. Because there are large differences in the types of conditions treated by urban and rural hospitals, it raises a concern that the case mix is driving the differences vs. actual differences in adjusted resource use.
  - Comment #3261: We thank NQF for the opportunity to comment on the cost and resource use measure, #2158 Medicare Spending per Beneficiary (MSPB). We have several areas of concern with this measure that should be addressed prior to endorsement. In the MSPB results that have been published by the measure developer, there is a noticeable clustering of large, urban, teaching hospitals that treat a large proportion of low-income patients with higher MSPB index rates than their community hospital counterparts. We believe this is due to insufficient severity adjustment in the measure that does not account for the ranges of patient complexity that exist between and within MS-DRGs. Since large, urban, teaching hospitals, with a large share of low income patients have the capability to treat more complex patients, whereas community hospitals often do not, they have a higher proportion of complex cases that require more hospital resources and are also more likely to have home care or skilled nursing care following the inpatient admission. This mix of more complex patients could be a contributing factor to the clustering being seen in the results. We would expect a more normal distribution of MSPB results across all hospitals if the measure were appropriately severity/risk adjusted and adjusted for outliers.
- Risk stratification using MDC criteria alone is inadequate and will introduce significant variability in the MSPB rating based upon patient-specific and diagnosis-specific factors that are not adequately encompassed in the MDC classification.
  - Comment #3271: NASS is concerned that the risk stratification for MSPB does not have adequate granularity to differentiate significant cost drivers. Specifically, the proposal to stratify cases by major diagnostic category (MDC).There is significant evidence that MDC classification does not

#2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB)
accurately encompass the factors that contribute to cost of care, and there are significant inaccuracies in the administrative data that contributes to the MDC Classification. Risk stratification using MDC criteria alone is inadequate and will introduce significant variability in the MSPB rating based upon patient-specific and diagnosis-specific factors that are not adequately encompassed in the MDC classification. A more comprehensive classification that includes an algorithm that includes CPT codes and procedure specific information would be more useful than a stratification based upon MDC alone. o Concern that the 90-day look-back period to capture a patient's comorbidities in order to determine the HCC score is insufficient. **Steering Committee Response:** The Committee reviewed the concern around the clustering of large, urban, teaching hospitals that treat a large proportion of low-income patients raised by the commenters. They voiced concerns on both sides of the issue of including SES adjustment with some members agreeing with the Commenter that disadvantaged patients with multiple complex conditions will require more resources to treat, while other members argued that including SES variables in the risk adjustment model will mask disparities in cost performance among different groups of patients. The Committee acknowledged that hospitals are legitimately held accountable for taking appropriate care of patients within the case mixes and making sure that these patients receive appropriate post-acute care, however, the availability of these support services will vary from community to community. The Committee recommended that additional work be considered in this area, specifically the appropriateness of including dual-eligibility in risk adjustment models for resource use measurement. The Committee considered the major diagnostic category (MDC) risk stratification criteria, specifically applying the risk adjustment within MDC to be generally appropriate for this application. For the purposes of performance measurement, factors that are included in the risk adjustment model should be present at the start of care – thus including procedure codes that occur during the measurement period would not be appropriate. The Committee also expressed concern over the 90-day look back period but ultimately agreed that the performance of the models did have a slightly improved model fit over the models with a year of look-back.
Final Steering Committee Recommendation for Endorsement: Y-17; N-8

Measure Not Recommended

#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries

Steward: Centers for Medicare & Medicaid Services

Description: The Payment-Standardized Total Per Capita Cost Measure for Medicare FFS Beneficiaries assesses the per capita (per beneficiary) cost of health care services for Medicare FFS beneficiaries enrolled in Parts A and B and attributed to medical group practices. The measure includes all Medicare Part A and Part B costs during a calendar year and is payment-standardized and risk-adjusted (using patient demographics and medical conditions) to account for any potential differences in costs among providers that result from circumstances beyond the physician's control. Under CMS' attribution rule, beneficiaries are attributed on the basis of the plurality of primary care services, to those medical group practices with the greatest potential to influence the quality and cost of care delivered to Medicare FFS beneficiaries.

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

Data Source: Administrative claims

Level of Analysis: Clinician : Group/Practice

Costing Method: Standardized pricing

Target Population: Senior Care

Resource Use Service Categories: Inpatient services: Inpatient facility services; Inpatient services: Evaluation and management; Inpatient services: Procedures and surgeries; Inpatient services: Imaging and diagnostic; Inpatient services: Lab services; Inpatient services: Admissions/discharges; Inpatient services: Labor (hours, FTE, etc.); Ambulatory services: Outpatient facility services; Ambulatory services: Emergency Department; Ambulatory services: Evaluation and management; Ambulatory services: Procedures and surgeries; Ambulatory services: Imaging and diagnostic; Ambulatory services: Lab services; Durable Medical Equipment (DME); Other services not listed

STEERING COMMITTEE MEETING [May 8-9, 2013]

Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. High Impact: 1b. Performance Gap, 1c. Measure Intent)

1a. Impact: **H-20**; **M-2**; **L-2**; **I-0**; 1b. Performance Gap: **H-11**; **M-10**; **L-4**; **I-0**; 1c. Measure Intent: **H-8**; **M-13**; **L-4**; **I-0**; 1. Overall: **H-11**; **M-10**; **L-4**; **I-0**

<u>Rationale:</u> While evaluating the measure's importance to measure and report, the Committee agreed that the subcriteria were met and provided the following rationale:

- There was general agreement that this represents a high impact area of healthcare.
- The Committee was concerned, however, that the results of the measure may not be actionable

#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries

because of the attribution method.

- The measure does not present a consistent breakdown of disparities (race, dual eligible status, etc.).
- The inclusions of pharmacy costs would present a more accurate picture of costs.
- The measure applies to 7,000 groups across the country and covers 75% of physicians. This represents a majority of physicians but a minority of groups. This measure would therefore benefit large groups with a value modifier.
    - After the Steering Committee meeting, the developer clarified that the measure covers 45% of physicians, not 75% as stated during the in-person meeting.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

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

2a. Reliability: **H-5**; **M-18**; **L-1**; **I-0**; 2b. Validity: **H-0**; **M-13**; **L-12**; **I-0**

Rationale: While evaluating the measure's scientific acceptability, the Committee agreed that the subcriteria were met and identified 3 major issues:

1) Attribution Method
2) Risk Adjustment Model
3) Exclusions

1) Attribution Method
    - The Committee was concerned about the general construction of the attribution approach.
    - Stage 1 of the attribution model assigns patients to physician groups by looking at number of visits with a primary care physician. The first stage of the attribution model does not consider the number of visits with Physician Assistants (PA) or Nurse Practitioners (NP). The lack of consideration of PA and NP visits was questioned, as PAs and NPs increasingly deliver more primary care.
    - The developer responded that this measure was designed according to requirements in statue to capture per capita costs for services delivered by physicians; thus physicians serve as the entry point to the attribution model.

2) Risk Adjustment Model
    - The Committee expressed concerns about the inclusion of dual-eligible status and gender in the risk adjustment model. The developer responded that the model was originally developed for the Medicare Advantage program and not necessarily for this measure.
    - Questions arose over whether the inclusion of SES and demographic factors could obscure the identification of disparities in care.
    - The Committee found the Hierarchical Condition Category (HCC) risk adjustment methodology with demographic factor adjustments to be weak in this application.

3) Exclusions
    - The Committee questioned the exclusion of deaths, part-year beneficiaries, and Medicare Advantage beneficiaries; these areas represent significant opportunities for improvement in

#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries

reducing spending.

- The Committee was concerned that excluding patients with Medicare Advantage presented a significant opportunity for "gaming" of the measure. High cost patients could be shifted to Medicare Advantage to prevent costs from being captured and attributed to the practice.
- The developer stated that some physician stakeholders did not agree that the inclusion of part-year beneficiaries was a fair representation of the cost of care for their patients.

Other issues

- The developer calculated a reliability score by measuring the between medical group variance compared to within medical group variance. The Committee expressed concern regarding these reliability testing results which showed that for medical group practices with at least 25 EPs and 20 attributed beneficiaries, the average reliability was 0.95, and 99 percent of groups had a reliability exceeding 0.50, and 96 percent of groups had a reliability exceeding 0.70.
- The Committee was concerned that the use of Tax ID numbers (TIN) may not be an accurate way to identify physician groups. Several small groups may bill under the same TIN giving the appearance of a larger group for the purposes of this measure. The developer agreed that this may be a legitimate concern; however, because the TIN is the unit of payment, it is still a legitimate method to aggregate costs.

3. Feasibility: H-19; **M-5**; **L-1**; **I-0**

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

Rationale: While evaluating the measure's feasibility, the Committee agreed that the subcriteria were met and provided the following rationale:

- The data for the measure is being collected and is a byproduct of the care process.
- Data is generated electronically
- Providers are not able to implement this measure without CMS. Commitment must be made from those with the data to make it publicly available.

4. Usability: H-4; M-14; L-7; 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: While evaluating the measure's usability, the Committee agreed that the subcriteria were met and provided the following rationale:

- The Committee largely agreed that the measure will be most useful when paired with quality outcome measures.
- This measure can drive change by placing primary care physicians as the responsible entity.
- Groups are in the best position to impact coordination of care and affect the access of care by the individual.

#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries

- Significant variation within groups can be masked by group-level reporting; physician-level reporting would eliminate that masking, but presents its own challenges.
- Consumers/Purchasers would find physician-level reporting to be the most actionable.

5. Related and Competing Measures

Potential harmonization issues relating to #1598 Total Resource Use Population-based PMPM Index (HealthPartners) were discussed by the Committee:

- The Committee reviewed areas of conceptual and technical similarities and differences between the two measures, noting that both measures are per capita, non-condition specific and capture standardized prices; however, the measures address different but overlapping target populations. NQF#1598 addresses the commercially insured population, and NQF#2165 addresses the Medicare population.
- The Committee considered whether the differences in target population and the differences in approach to standardization of prices were sufficient to justify recommending that the two measures not be harmonized and remain distinct. As part of this discussion, the Committee considered the potential value and burden for this; specifically, whether the differences in the technical specifications are necessary, affect interpretability across the measures, or affect data collection burden.
- The Committee reviewed the key differences between the measures and agreed that there was little room for increased alignment between the measures given the unique characteristics of the two target populations and measure intent. The Committee stated that the differences in the data sources resulting from the differences in the target populations for the two measures drive the differences in the technical specifications for the measures.
- Some members of the Committee suggested that the developers consider potential harmonization of their attribution approach. They discussed that providers could better interpret how their patients are assigned to them if the attribution approach is similar for their Medicare and commercial patients.
- The Committee also discussed differences of pharmacy data; the HealthPartners measure does include pharmacy data when available and the CMS measure does not. Members of the Committee recommended that CMS consider experience from commercial payers in handling missing pharmacy data.

Steering Committee Recommendation for Endorsement: **Y-11**; **N-14**

Rationale:

- The Committee was concerned about the construction of the measure and the ability of the attribution approach to capture costs appropriately and assign them to appropriate providers.
- The exclusion of Medicare Advantage, part-year beneficiaries, Part D, and deaths limit the utility of the measure to address high-cost, high-priority areas of healthcare.
- Reporting at the group level may not provide actionable information and mask significant intra-group variation.
- The Committee did not reach consensus on this measure. The Committee considered this vote "preliminary" and will likely reconsider after the developer's responses and public and member comments have been reviewed and discussed.

Public and Member Comment [July 9 – August 7, 2013]

This measure received comments from eighteen organizations/ individuals. Several commenters

shared support for the concept and intent of the measure, urging CMS to make revisions to the attribution approach, risk adjustment algorithm, reliability and validity of the measure and to bring the measure back to NQF for endorsement, as this is an area where measures are needed and would provide insight into the costs of healthcare to Medicare. In addition to the support for the concept and intent of the measure, one commenter also acknowledged "provider concerns over the attribution of total cost of care to primary care physicians who are not part of an organized health system. But purchasers have come to expect care to be coordinated among providers and see the need to incentivize such coordination. Moreover, measures such as this one will help primary care physicians to understand the cost implications of their referral recommendations."

Steering Committee Response:
The Steering Committee raised several concerns with the construct of the measure, which the developer has been working to analyze and address during the past few months. Responses to the committee's concerns and additional analysis performed by the developer were shared with the committee on their August 28th call. The Steering Committee had the opportunity to review all comments and the developer's analysis and re-affirmed their decision to not recommend the measure for endorsement.

NQF will work with the developers to determine when it can next be reviewed for endorsement. This would happen when the next Cost and Resource Use project is scheduled.

The Steering Committee unanimously agreed that cost and resource use measures must be paired with quality measures in order to understand and make decisions about care. The Committee agrees with the commenters that measures of efficiency, and ultimately value are critical tools needed to improve the efficiency of US health care system, specifically encouraging shared accountability and team-based care.

The Steering Committee acknowledged the consumer perspective that care should be coordinated among providers; however, the Steering Committee was split over the idea that it may be inappropriate to hold primary care providers accountable for the cost of care provided to patients by other specialists, through inpatient care or through post-acute care, as primary care providers have limited ability to control these costs. In the current state of care delivery, health care is accessed in many ways. Many patients select their own primary and specialty care physicians, making decisions to see providers on their own, without coordination with their PCP or PCP group. Several Committee members stated that this may be appropriate in markets with integrated care delivery networks or where patients identify with a PCP or PCP group voluntarily or by assignment; however, in the current fragmented state of care delivery this attribution approach is not preferred Several other committee members stated that this level of accountability for providers is the entire rationale for the measure and should help push providers to be better organized to reduce costs.

The remaining comments addressed several themes listed below, along with a description of the comments received.

#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries

Attribution

- Several commenters agreed with the Committee that primary care physicians or specialists who may be attributed patients because they provided primary care services to that patient have limited ability to control the cost of care provided to patients by other specialists, through inpatient care or through post-acute care. The majority of commenters agreed that it may be inappropriate to hold these providers accountable for these costs of care. The commenters also agreed that this may be appropriate in markets with integrated care delivery networks; however, in the current fragmented state of care delivery this attribution approach is not supported.
- Additionally, several commenters shared the Steering Committee concerns that patients and their associated costs may potentially be attributed to specialists who provide primary care services that are Medicare allowable charges and questioned the appropriateness of this.
- Several commenters shared the Steering Committee concern that visits with non-physician providers (PAs and NPs) are not taken into account in the attribution model until the second stage, as non-physician providers are increasingly delivering more primary care.
- Given the various concerns about the attribution approach, several commenters called into question the reliability and validity of the measure, noting the Steering Committee's split vote as to whether the measure was in fact valid.

Steering Committee Response:

The Steering Committee acknowledged many of the same concerns with the attribution approach. The Steering Committee stated concern that patients and their associated costs may potentially be attributed to specialists who provide primary care services that are Medicare allowable charges. This is particularly significant in the case of patients who receive long-term care for chronic conditions, who may receive many primary care services from specialists treating them for their chronic conditions, who are then attributed to a medical group practice based on the plurality of Medicare allowable charges. The Committee noted the distinction that specialists can provide primary care services through visits other than primary care visits.

The Committee was ultimately split on the concern that physicians have little ability to control the cost of care provided to patients by other specialists, through either inpatient care or post-acute care. Several Steering Committee members raised concern that it may be inappropriate to hold these providers accountable for these costs of care. Further, several Committee members stated that this may be appropriate in markets with integrated care delivery networks; however, in the current fragmented state of care delivery this attribution approach is not preferred. On the other hand, several other committee members stated that this level of accountability for providers is the entire rationale for the measure and should help push providers to be better organized to reduce costs.

The Steering Committee agreed with commenters that there are issues with both the first and second stage of the attribution approach. In the first stage, visits with non-physician providers (PAs

and NPs) are not taken into account in the attribution model until the second stage, as non-physician providers are increasingly delivering more primary care. The Committee strongly encourages CMS to include non-physician providers in the first stage of the attribution approach. Further, primary care services as defined by this measure may not always represent actual primary care visits by primary care providers. The Committee encourages CMS to update this attribution approach.

Exclusions

- One commenter expressed concern that the exclusions of death and Medicare Advantage beneficiaries impact the usability of the measure.
- One commenter expressed concern that Medicare Part D (prescription medications) was excluded from the measure.

Steering Committee Response:
The committee was split on the reliability and validity of this measure but ultimately agreed that a number of issues, including the exclusions of death needed to be addressed before recommending this measure for endorsement. Additionally, Medicare Part D payment is an important area for measurement and improvement. CMS should consider approaches to including this data for beneficiaries with Part D coverage.

Reliability

- One commenter requested that the measure developer not publically report results for any provider group with reliability scores less than 0.70.
- One commenter stated that the measure is only reliable for groups of 25 or more eligible professionals; however, nearly half of all Medicare physicians practice in groups of fewer than 10 eligible professionals. As the measure will be used as part of CMS' value-based modifier calculation, the commenter questioned how this will impact smaller physician groups and solo practitioners.

Steering Committee Response:
While NQF does not require a specific cut-off for reliability testing, the Committee does encourage CMS to report information on provider groups that have adequate reliability in performance score and sample size. This measure should only be used for 25 or more eligible professions since this is the scope of measure testing.

Risk Adjustment

- One commenter expressed concern that the risk adjustment model might not adequately capture the differences in patient population for different specialties, particularly those who treat patients with uncommon and very severe diseases.
    - Comment #3253: We appreciate the opportunity to comment on the measure #2165 (Payment-Standardized Total Per Capita cost Measure for Medicare Fee-for-Service Beneficiaries). We have some concerns about this measure and its potential for use as a component of the value-based modifier. From the measure description

#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries

and information provided, it is unclear how this measure would be applied. We are concerned about the broad nature of this measure and the fact that it looks across different specialties rather than within each specialty. We understand that the risk adjustment takes into account complexity of disease; however, we are concerned that the risk adjustment model might not adequately capture the differences in the patient population for different specialties. We are concerned that certain specialties, particularly cognitive specialists like rheumatology caring for patients with uncommon and very severe diseases, as a whole might fare worse than others if this measure is applied across specialties. In addition, we reviewed the risk adjustment model and do not believe it adequately captures the scope and complexity of conditions that rheumatologists care for. The exclusion of consideration of specific patient populations in the risk adjustment model would put providers or centers who treat a large number of these patients at a disadvantage. We would urge any assessment of providers for efficiency to look within a specialty rather than across specialties and that the risk adjustment model be thoroughly reviewed through specialty societies.

- Several commenters stated that the HCC model, which was developed for the Medicare Advantage program, does not adequately account for risk for purposes of analyzing physician group resource use, as it was designed to risk adjust large patient populations for insurance rate determination.

Steering Committee Response:
The Committee generally agreed that while this HCC-risk adjustment model was developed for Medicare Advantage it was appropriate but weak in this application. The HCC model does not include as many diagnostic categories as many commercially available risk adjustment models and therefore may not be as accurate in assigning the appropriate risk categories for rare conditions. However, given the broad use of HCCs across Medicare programs the Committee agreed that this approach was sufficient for this application.

Final Steering Committee Recommendation for Endorsement: Y-12; N-13