TO: Consensus Standards Approval Committee (CSAC)

FR: Taroon Amin, Ashlie Wilbon, and Lindsey Tighe

RE: Cost and Resource Use Project Update

DA: July 5, 2013

The CSAC will review the project status of the Cost and Resource Use project at its July 10th-11th in-person meeting.

This memo includes a summary of the project, recommended measures as of the May 8-9th Steering Committee in-person meeting, and themes identified from the Cost and Resource Use Steering Committee discussions.

This project is following the National Quality Forum's (NQF) version 1.9 of the Consensus Development Process (CDP). Member commenting on the Cost and Resource Use measures runs from July 9th through August 7th, 2013.

Accompanying this memo are the following documents:

1. **Cost and Resource Use Draft Report for Commenting.** The draft report reflects the Steering Committee discussions and recommendations from their May 8-9th in-person meeting.

CSAC ACTION REQUIRED
Review the current status of the measures being reviewed in the project and provide input on Steering Committee discussions and overarching issues.

BACKGROUND

This project focuses on evaluating and endorsing cost and resource use measures. At this point, the project entails evaluation of non-condition specific measures of total cost, using per-capita or per-hospitalization approaches. NQF anticipates that additional phases of this project may be necessary to evaluate condition-focused measures, depending on availability of funding, so individuals nominated may be asked to serve on a second phase, as outlined below:

Phase One: Non-condition specific per capita or per hospitalization measures

Phase Two: Condition-specific per capita and condition-specific episodes beginning with the following condition areas:

- Cycle 1: Cardiovascular
- Cycle 2: Pulmonary, Diabetes

DRAFT REPORT

On May 8-9, 2013 the Cost and Resource Use Steering Committee evaluated two new measures against NQF's Resource Use Measure Evaluation Criteria:

- 2158: Payment-Standardized Medicare Spending Per Beneficiary (MSPB)
- 2165: Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries

The report was posted for comment on July 8th. The measure developers both provided memos to be posted with the draft report addressing concerns raised by the Steering Committee during the May 8-9 in-person meeting. These memos have been appended to the draft report.

COST AND RESOURCE USE SUMMARY

	MAINTENANCE	NEW	TOTAL
Measures under consideration	0	2	2
Measures withdrawn from consideration	0	0	0
Measures Recommended	0	1	1
Not recommended	0	1	1

NQF#2158: Payment-Standardized Medicare Spending Per Beneficiary (MSPB) was recommended for endorsement; NQF#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries was not recommended for endorsement. Please reference the draft report for the full discussion of measure specific issues and the Steering Committee voting results on each of the major criteria and the overall endorsement decision.

Overarching Issues for CSAC Discussion

During the Steering Committee's discussion of the measures, several overarching issues emerged that were factored into the Committee's ratings and recommendations for multiple measures:

1. *Appropriateness of inclusion of markers of socioeconomic status (SES) in risk adjustment models*

The NQF guidance supporting the scientific acceptability criteria for risk adjustment (2b.4) indicates that risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status (SES), or gender. NQF recommends that measures be stratified by race and socioeconomic status rather than adjusting away differences that may be due to disparities in the quality of care provided. Given this guidance, the Committee discussed the appropriateness of including markers of socioeconomic status in the risk model at length in regards to both measures, each with a different approach to accounting for these differences.

- NQF#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries includes gender and dual eligibility status in the version of the hierarchical condition category (HCC) risk adjustment model used in the measure.

NQF#2158: Payment-Standardized Medicare Spending Per Beneficiary (MSPB) does not include adjustments for dual-eligibility.

2. Risk Adjustment

In response to requests from various stakeholders and NQF members, NQF was asked to consider the implications of endorsing a single cost/resource use measure that has been tested with multiple risk adjustors. This would enable the measure to be used interchangeably with different risk adjustors based on user need. The need for flexibility in risk adjustors is reflective of the healthcare market in which different regions and healthcare systems have invested in a single risk adjustor that may not be one that was used in an endorsed measure of interest.

While allowing flexibility would potentially enable markets and users to continue with the risk adjustment model that they have purchased and already have in place, to facilitate national comparisons a single tool must be used. Even with acceptable testing of the measure with each risk adjustor individually, comparability among the measures using the different adjustors is limited. Further, in order to ensure that measure users were not inadvertently comparing measure results from the measure using a different risk adjustor, each of the measure-risk adjustor combinations would need to be endorsed separately and have different endorsement numbers. As such, the measures would arguably be competing measures, with the same measure types, the same approach to measurement, and targeting the same population.

In response to the various concerns raised by this issue, the Committee also considered the state of resource use measure development and commercially available risk adjustment methodologies determining that:

1. In order to be useful in making valid conclusions about performance, particularly relative performance, all entities should be measured in the same manner.
2. If a measure is submitted using multiple risk adjustment models, the developers must submit empirical analyses to demonstrate comparability of measure results. These analyses should compare the same patients, distribution of diseases, distribution of cost, as well as measure and compare different risk adjusters. The results should analyze both the differences in relative ranking of providers and the differences in the performance measure results.

Cost and Resource Use 2012

DRAFT TECHNICAL REPORT FOR REVIEW

July 9, 2013

NATIONAL QUALITY FORUM

Comments due by August 7, 2013 by 6:00 PM ET.

Contents

Introduction 3
Measure Evaluation 4
    Overarching Issues 5
    Measure Specific Issues 7
Recommendations for Future Measure Development 8
Next Steps 9
Measure Evaluation Summary 14
    Recommended Measure 15
    Measure Not Recommended 19
Notes 23
Appendix A: Measure Specifications 24
Appendix B: Project Steering Committee and NQF Staff 28
Appendix C: Measures Endorsed in Cost and Resource Use Since April 2012 30
Appendix D: Related and Competing Measures 31

Cost and Resource Use 2012

DRAFT TECHNICAL REPORT

Introduction

Per capita healthcare spending in the United States is unmatched by any country in the world.[1] This high rate of spending, however, has not resulted in better health for Americans. In fact, higher spending has not decreased mortality, increased patient satisfaction, or led to improvements in access or higher quality of care.[2,3,4] This phenomenon of high spending with variable outcomes points to a system laden with waste. The contributing factors to this concerning trend are as complex as the healthcare system itself, with physician practice patterns, regional market influences, and access to care as major drivers. Meanwhile, the United States' healthcare spending continues to increase at a rate of seven percent per year and is largely focused on treating acute and chronic illness rather than preventive care.[5] By improving efficiency, there is potential to reduce the rate of cost growth and improve the quality of care provided simultaneously. Evidence shows that not all care leads to better outcomes; thus, some portion of these current costs may be unnecessary. To identify and provide incentives for providers to deliver high quality, lower-cost care requires quality and resource use measures.

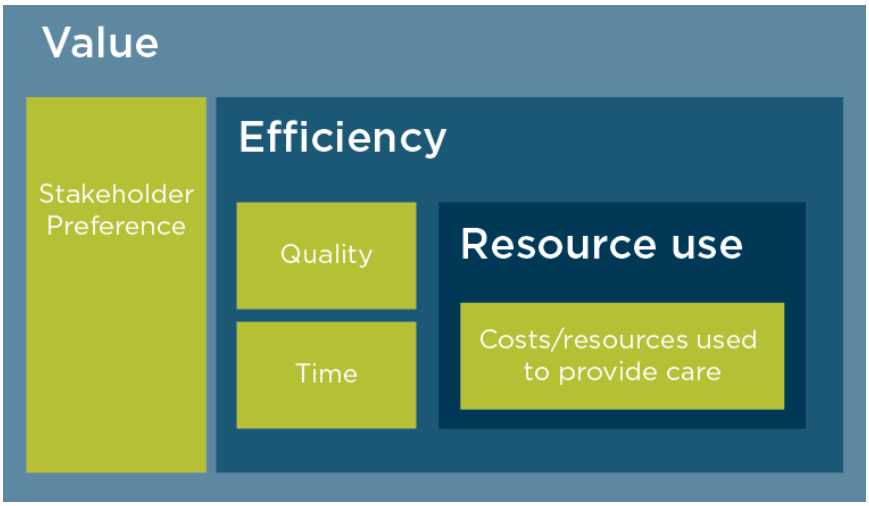
The National Quality Strategy's (NQS) three aims—better care, affordable care, and healthy people, healthy communities—have intensified the need to identify measures that address cost and align them with the relevant quality measures already in the marketplace. The NQS specifically identifies affordability as a target area for improvement, with goals of:

1) Ensuring affordable and accessible high quality health care for people, families, employers, and governments.
2) Supporting and enabling communities to ensure accessible, high quality care while reducing waste and fraud.

As ongoing health reform efforts focus on expanding coverage, increasing access to care, and reducing costs, it is important to understand how resources are currently being used in the system in the context of quality, preferably related to health outcomes. Aligning resource use (or cost) and quality measures will enable the system to better evaluate efficiency of care. Several provisions in the Affordable Care Act (ACA), slated to be implemented over the next three years, require using resource use data to further support efforts to move toward a value-based purchasing (VBP) payment model. Resource use data will be included on the physician compare website, as well as a physician value modifier that will be used to adjust fee-for-service (FFS) payments by combining physician performance on quality and resources use.

In January 2010, NQF released the *Measurement Framework: Evaluating Efficiency Across Patient-Focused Episodes of Care,* which addressed cost and resource use as one of the three overarching domains for assessing efficiency. This framework advised that measures of resource use and cost should incorporate approaches to measure actual prices paid to providers, standardized prices, in addition to overall utilization. Further, inappropriate care, including failing to provide an evidence-based intervention to an eligible patient or administering an intervention that is unwarranted, cannot be efficient.

NQF's work around endorsing cost and resource use measures has built on this concept within the Efficiency Framework report that measures of cost and quality must be aligned in order to truly understand efficiency and value (Figure 1). NQF has defined efficiency broadly as the resource use (or cost) associated with a specific level of performance with respect to the other five Institute of Medicine (IOM) aims of quality: safety, timeliness, effectiveness, equity, and patient-centeredness.



To expand the NQF portfolio of endorsed cost and resource use measures that in turn could be used as building blocks toward understanding efficiency and value, in 2010 NQF embarked on its first effort to evaluate and endorse cost and resource use measures. Laying the foundation for NQF's current work, the definition of cost and resource use measures that are guiding this work and the scope of measures was broadly applicable and comparable measures of health services counts (in terms of units or dollars) that are applied to a population or event (broadly defined to include diagnoses, procedures, or encounters). A resource use measure counts the frequency of defined health system resources; some may further apply a dollar amount (e.g., allowable charges, paid amounts, or standardized prices) to each unit of resource use. This learning was captured in the final and technical reports, yielded the first eight endorsed cost and resource use measures in the NQF portfolio, and the NQF Resource Use Measure Evaluation Criteria. The work in this first consensus development project on cost and resource use measures serves as the foundation for this project.

Measure Evaluation

On May 8-9, 2013 the Cost and Resource Use Steering Committee evaluated two new measures against NQF's Resource Use Measure Evaluation Criteria:

- 2158: Medicare Spending Per Beneficiary measure (CMS)
- 2165: Total Per Capita Cost Measure for Medicare Beneficiaries measure (CMS)

To facilitate the evaluation, each of the committee members completed preliminary evaluation of the measures prior to consideration by the entire Steering Committee at the in person meeting. The Committee's discussion and ratings of the criteria are summarized in the evaluation tables beginning on page 11.

COST AND RESOURCE USE SUMMARY

	MAINTENANCE	NEW	TOTAL
Measures under consideration	0	2	2
Measures withdrawn from consideration	0	0	0
Measures Recommended	0	1	1
Not recommended	0	1	1

Overarching Issues

During the Steering Committee's discussion of the measures, several overarching issues emerged that were factored into the Committee's ratings and recommendations for multiple measures:

Risk adjustment

Socioeconomic Status

The NQF guidance supporting the scientific acceptability criteria for risk adjustment (2b.4) indicates that risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status (SES), or gender. NQF recommends that measures be stratified by race and socioeconomic status rather than adjusting away differences that may be due to disparities in the quality of care provided. Given this guidance, the Committee discussed the appropriateness of including markers of socioeconomic status in the risk model at length in regards to both measures, each with a different approach to accounting for these differences.

During the evaluation of Total Per Capita Cost Measure for Medicare Beneficiaries measure (NQF# 2165), the developers described that gender and dual eligibility status are both included in the version of the hierarchical condition category (HCC) risk adjustment model used in this measure. In the importance section of the measure submission, the developers indicated that there is data demonstrating disparities by population group, specifically dual eligible, noting that Medicare spending on dual eligible beneficiaries was almost two times higher than spending on non-dual eligible beneficiaries in 2008[6]. Therefore, including a marker of dual- eligibility in the measure's risk adjustment model has the potential of masking this difference.

Given that this modified HCC risk adjustment model was originally developed for Medicare Advantage plans, these demographic factors have been used historically by CMS actuaries to determine payments to these plans. The original intent of the model is to avoid risk selection of patients based on gender

and dual-eligibility status. The Committee was concerned that given the intended use of the original model it may not be suited for performance measurement where these factors are preferably excluded from risk adjustment to determine whether disparities in care exist.

In the evaluation of the Medicare Spending Per Beneficiary measure (NQF #2158), the measure developers did not include adjustments for dual-eligibility. While this measure also uses a version of the HCC risk adjustment model, the developers explained that they tested an exclusion of dual-eligible beneficiaries, as well as an inclusion of dual-eligible beneficiaries as a risk adjuster, but these adjustments did not result in major differences in the measure performance. Thus, the decision was made for the measure to include dually-eligible beneficiaries in the measure population, but not to include a dually-eligible risk adjuster in their version of the HCC risk adjustment model.

The Committee ultimately agreed that more guidance in this area is needed, particularly using markers of SES variables in outcome and resource use measures.

Look back period

In the evaluation of the Medicare Spending Per Beneficiary measure (NQF #2158), some Committee members were concerned with the application of the HCC risk adjustment to capture and identify pre-existing conditions. While the measure identifies these pre-existing conditions by looking back to conditions present in the 90 days prior to admission, the HCC model is designed for a full 12-month look back period. In response to this concern, the developers described their testing of various look back periods and concluded that the 90-day look-back period offered marginally superior performance to the 12-month look back period. This is possibly because the conditions that occurred closer to the hospitalization were more relevant to predicting patient severity. The Committee was ultimately satisfied with the response from the developer and the level of testing conducted to justify the 90 day look back period used in the measure.

Exclusion of Deaths

In both measures, patients who died were excluded from the measurement period. This decision was made based on testing showing that this subset of patients has a bi-modal distribution of costs caused by a significant number of patients who are high cost and a significant number of patients who are low cost. This distortion in the distribution of data may limit the validity of the model to predict costs for this subset of patients, compared to predicting costs of patients within a normal distribution. The high cost group likely represented those beneficiaries that received high intensity end-of –life care and died toward the end of the measurement period. Conversely, those that died earlier in the episode would show as low cost as they likely used resources for a shorter period of time. The Committee generally disagreed with this exclusion arguing that end of life care is a high-cost area for the Medicare program and is important for measurement and improvement. Further, excluding patients who die during the measurement period may create unintended negative consequences. For example, hospitals that provide intense care keeping patients alive will appear more costly than hospitals providing equally intense care but ultimately resulting in a patient death because the costs associated with the death have been excluded. Further, in the evaluation of Medicare Spending Per Beneficiary measure (NQF #2158), the Committee questioned the developers on the appropriateness of including hospice costs when deaths are excluded from the measure. This seems counterintuitive as patients entering hospice are expected to die, and thus costs associated with hospice care may be excluded from the measure.

Measure Specific Issues

During the Steering Committee's discussion of the measures, several issues specific to individual measures emerged.

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

The Committee noted several issues in their discussion of this measure, all linked to validity. The measure's attribution approach and the exclusion of pharmacy costs and Medicare Advantage patients were of specific concern.

Attribution

The attribution methodology chosen for NQF #2165 was strongly questioned by many members of the Steering Committee. While this attribution approach has been used in other CMS programs, including the Physician Group Practice (PGP) demonstration and more recently in the group practice reporting option (GPRO) of the Physician Quality Reporting System (PQRS), the Committee expressed concerned about a number of factors. First, the measure includes a two-step attribution rule in which the first step attributes beneficiaries to a medical group with affiliated primary care physicians (PCPs) whose services account for the largest amount of Medicare allowable charges within the measurement period. If the beneficiary is not assigned in the first step, they are assigned to any medical group in which they have seen at least one physician in the group, regardless of specialty, who has provided primary care services. Attribution to the medical group is based on which medical group provided the largest amount of Medicare allowable charges during the measurement period (inclusive of the charges by specialists physicians, nurse practitioners, physician assistants, and clinical nurse specialists). The attribution methodology assigns all health care services and associated costs for the beneficiaries to the medical group identified in either step one or step two of the attribution rule.

Members of the Committee expressed strong reservation that visits with non-physician primary care providers, specifically nurse practitioners and physician assistants, are not included as eligible visits for attribution to a provider group in the first stage. The developers described that this approach was guided by statute and is based on CMS' goal to provide feedback on resource use to physicians, but agreed that the inclusion of these providers should be considered in future iterations of this measure given their growing role in primary care.

Some members of the Steering Committee were also concerned that the attribution method limited the utility of the measure to improve cost performance for two primary reasons. First, even with the reports provided for this measure, there is limited information on the within group variation of costs. Feedback given to provider groups is rolled up to the group level and thus individual provider cost variation may be masked. Second, primary care providers may have limited ability to influence the cost of specialists, inpatient care and post-acute care and may be ultimately held responsible for these costs. In markets with integrated care delivery networks, this may be expected of primary care; however, the current fragmented state of care delivery does not support this attribution approach.

Due to the many concerns expressed about this attribution approach, many members ultimately agreed that the approach significantly impacted the validity of the measure and the accountability for costs should be explored differently to allow for shared accountability across providers.

Outpatient Pharmacy costs

Given that not all Medicare beneficiaries have Medicare Part D coverage, the measure does not include outpatient pharmacy costs. The Committee 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. They argued that since more than half of Medicare beneficiaries have Part D coverage, it is important to understand the cost drivers and variation in drug utilization and costs. Members suggested that the developer consider aggregating and reporting the measure by those who have Medicare Part D coverage and those who do not.

Exclusion of Medicare Advantage

Committee members were concerned about the exclusion of Medicare Advantage patients from the measure. Medicare Advantage Plans, also known as Medicare Part C, are health plans offered by private companies approved by Medicare. Members of the Committee argued that measuring cost for beneficiaries in the Medicare Advantage plans is equally as important as the Medicare fee-for-service population. The measure developers argued that it is often difficult to obtain utilization data for beneficiaries enrolled in Medicare Advantage Plans; as such, these beneficiaries were not be included in the measure. There was concern that there could be gaming with large, sophisticated practices encouraging higher cost fee-for-service beneficiaries to switch to Medicare Advantage plans, enabling the practice to continue seeing these patients without inclusion of their costs in the measure.

The Committee requested that the developers reconsider this exclusion. Many members agreed that this issue would not ultimately influence the final endorsement recommendation but does challenge the validity of the measure as constructed.

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

Exclusion of Transfers

In this measure all beneficiaries that 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 that 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.

Recommendations for Future Measure Development

During their discussions the Committee identified numerous areas where additional measure development is needed:

In order to understand efficiency, cost and resource use measures should be paired with:

- appropriateness/overuse measures

- outcome measures
- process measures
- clinical data and patient reported outcomes

Other gaps noted included:

- Measures capturing variations in cost and outcomes for potentially high cost patients (i.e. cardiovascular or diabetes patients)
- Episode-based cost and resource use measures
- Measures capturing actual prices paid to providers by health plans

Next Steps

Harmonization Discussion

The Steering Committee considered potential harmonization issues between the NQF#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries and the previously endorsed NQF#1598 the Total Resource Use Population Based PMPM measure developed by HealthPartners. A summary comparison of these measures is captured in Appendix D. The goal of this harmonization effort is to reduce measurement burden for providers and implementers, while improving interpretability for patients and facilitating alignment of measurement across public and private sector. In its preliminary recommendation at the in-person meeting, the Steering Committee did not recommend endorsement of Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries (#2165), but agreed to discuss areas of potential harmonization between the two measures.

The Steering Committee reviewed areas of conceptual (intent of the measure) and technical (how the intent of the measure is operationalized) similarities and differences between the two per capita measures. Prior to the Committee's discussion of potential harmonization areas between the two measures, NQF staff facilitated early discussions with the developers from each measure to identify possible areas of alignment. The developers were asked to submit a joint letter to the Committee outlining areas of potential alignment and key differences. Upon review, the Committee considered how the two measures would provide consistent measure results by improved interpretability across levels of analysis and data sources.

Similar resource use measures are defined as the following:

- same measure types (e.g. per episode, per capita),
- measure the same costs/resources (e.g. actual cost vs. standard prices, resource service categories),
- and address the same population (e.g. diabetic patients).

NQF#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries and NQF#1598 the Total Resource Use Population Based PMPM are both per capita, non-condition specific measures that capture standard prices; however, NQF#2165 addresses the Medicare population, and NQF#1598 addresses the commercially insured population. The measures are both risk adjusted; however, NQF#1598 the Total Resource Use Population Based PMPM uses a commercial risk

adjustment methodology developed and calibrated specifically for the commercially insured population (Johns Hopkins University's Adjusted Clinical Groups (ACG) Case Mix System whereas NQF#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries uses the CMS Hierarchical Condition Category (CMS-HCC) risk adjustment methodology designed for Medicare FFS beneficiaries.

While evaluating these similar resource use measures for potential harmonization, the Steering Committee was asked to assess both the value and the burden of recommending that the measures are not harmonized and remain distinct; specifically, whether the differences in the technical specifications are necessary, affect interpretability across the measures, or affect data collection burden. The Committee considered whether the measures had both sufficiently different populations and standardized costing approaches to justify the burden of having two similar measures. The Steering Committee was not asked to review the measures for harmonization of risk adjustment models, risk stratification approaches, and statistical methods for estimating measure results as this is not recommended under NQF guidance.

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, including risk adjustment methodologies. Given the different patient populations, the Committee discussed the challenges to align the risk adjustment methods and the payment standardization methodologies. 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 includes 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. For example, it may be possible to calculate the measure for people who have a pharmacy benefit and those without to create a blended per-member per-month measure result.

Risk Adjustment

In response to requests from various stakeholders and NQF members, NQF was asked to consider the implications of endorsing a single cost/resource use measure that has been tested with multiple risk adjustors. This would enable the measure to be used interchangeably with different risk adjustors based on user need. The need for flexibility in risk adjustors is reflective of the healthcare market in which different regions and healthcare systems have invested in a single risk adjustor that may not be one that was used in an endorsed measure of interest. In order to use the endorsed measure (including the risk adjustor), a potential user must weigh the benefits of the measure against the additional investment in another (proprietary) risk adjustor. This introduces a major barrier to the market and the uptake of endorsed measures as organizations often have limited resources; transitioning to another tool is financially inefficient as it introduces new licenses fees and opportunity costs.

While allowing flexibility would potentially enable markets and users to continue with the risk adjustment model that they have purchased and already have in place, to facilitate national

comparisons a single tool must be used. Even with acceptable testing of the measure with each risk adjustor individually, comparability among the measures using the different adjustors is limited. Further, in order to ensure that measure users were not inadvertently comparing measure results from the measure using a different risk adjustor, each of the measure-risk adjustor combinations would need to be endorsed separately and have different endorsement numbers.

Endorsing a single measure with multiple risk adjustors or separate measures (each with different adjustors) presents challenges in applying some of NQF's evaluation criteria and guiding principles for endorsement and national comparisons:

- NQF endorses national standards for performance measures that are intended for both accountability and performance improvement.
- In order to be useful to make conclusions about performance, especially relative performance, all entities need to be measured exactly the same way.
- NQF seeks to endorse the best from among competing measures whenever possible in order to minimize the confusion created when accountable entities are scored and ranked differently based on differences in measure specifications.
- The measure evaluation criteria must be applied to every measure submitted for endorsement. The following criteria are specifically challenging as it relates to multiple risk adjustors:
    - Scientific Acceptability Criteria 2b4 Risk Adjustment/Stratification for Outcome or Resource Use Measures. For resource use measures, an evidence-based risk adjustment strategy (e.g. risk model) is specified and is based on patient clinical factors that influence the measured outcome (but not disparities in care or the quality of care) and are present at the start of care
    - Scientific Acceptability Criteria 2b6 Comparability of Multiple Data Sources. If multiple data sources/methods (e.g. risk adjustment approaches) are specified, there is a demonstration that they produce comparable results.
    - Scientific Acceptability Criteria 2b2. Validity testing. NQF criteria allow testing of either the data elements or the measure score.

To facilitate discussion of this issue and fully solicit input from multiple stakeholders on this issue, the Resource Use Steering Committee was asked to provide input for consideration of NQF policy on this issue. In addition to the Committee, other stakeholders were invited to participate in this discussion including measure developers, statisticians, purchasers, and other measure users impacted by this issue. Specifically, a member of the Society of Actuaries (SOA) was invited to present to the Committee to discuss the implications of a 2007 SOA report, A Comparative Analysis of Claims-Based Tools for Health Risk Assessment. In this report, it was summarized that several risk adjustors evaluated in this study (e.g., ACG's, ETG's, DxCG's) had comparable performance[7].

Specifically related to the challenges this issue presents to current NQF policy and guidelines, the Committee was asked to consider and provide rationale for the following questions and potential options for future measure submissions:

- Should NQF consider changes to current policy on this issue and adopt one of the following :

    1. Endorse one measure, with one measure number, including multiple risk adjustment models for the user to pick from.

- Developers would be required to demonstrate comparability of the results with each adjuster.

2. Endorse multiple measures with different numbers, each measure with a different risk adjustment model.
    - Assume performance scores are not comparable
    - Each additional measure must be evaluated against criteria
    - All specifications, other than the risk model variables and coefficients should be identical
    - Requires justification of endorsement of multiple competing measures
3. Endorse one measure, with one measure number, with multiple risk adjustment models
    - Developer would not be required to demonstrate comparability.

In response to the various concerns raised by this issue, the Committee also considered the state of resource use measure development and commercially available risk adjustment methodologies determining that:

1. In order to be useful in making valid conclusions about performance, particularly relative performance, all entities should be measured in the same manner.
2. If a measure is submitted using multiple risk adjustment models, the developers must submit empirical analyses to demonstrate comparability of measure results. These analyses should compare the same patients, distribution of diseases, distribution of cost, as well as measure and compare different risk adjusters. The results should analyze both the differences in relative ranking of providers and the differences in the performance measure results.

More broadly, the Committee raised concern that there may be flaws of risk adjustment systems using claims or administrative data for resource use measurement. Patients who receive higher intensity treatment and thus generate more claims may potentially be assigned to a higher risk category or severity level, resulting in a higher expected cost. The Committee recommended monitoring of unintended negative consequences of this phenomenon.

Attribution

In cost measurement, attribution is the step in specifying measures that identifies is the responsible entity(s) for the performance results. While similar to the level of analysis, attribution specifically determines what proportion of the costs or resources are assigned to a single provider, divided amongst a group of providers, or some combination thereof. The level of analysis often aligns with the attribution approach, however, often focuses on the lowest level at which the costs can be rolled up and reported (e.g., physician, physician group, state, national). This is often dependent on measure testing to determine stability (or reliability) of the measure results with certain sample sizes.

While NQF seeks to endorse standardized performance measures intended for both accountability and performance improvement that can be used for national comparisons, users of cost and resource use measures often prefer flexibility in the attribution approach to accommodate specific applications, the unique attributes of their healthcare system or market and allow the opportunity to consider input from the attributable entities. Further, with no accepted gold standard for attribution or uniform guidance on best practices for attribution, it becomes difficult to determine how best to integrate it into the measure submission and evaluation process while trying to meet various needs. In response to the need for flexibility, under the direction of NQF's first Resource Use Steering Committee in 2010, the attribution

approach was allowed to be submitted as measure specifications or as optional guidelines for users to consider when implementing the measure.

In response to some confusion resulting from attempting to decipher specifications versus guidelines in the submission, the attribution approach was included on the submission form as guidelines only in the current resource use project. In both of the measure submissions submitted to this project, however, the attribution approach was included in the submission as specifications key to the implementation of the measure, and they were evaluated as such by the Committee. In an effort to continually evaluate and improve the NQF evaluation process of cost and resource use measures, the current Resource Use Committee was asked to reconsider the implications of requesting the attribution approach as guidelines or specifications. Specifically, they were asked to provide input on the potential for variation in measure results and impact on comparability of resources use measures that are implemented with various attribution approaches. Additionally, recognizing that quality measures also use attribution determine the responsible entity for performance results that is included in the specifications, the Committee was asked to consider whether resource use measures were sufficiently unique such that the attribution approach would not need to be specified and could continue to be submitted as guidelines enabling user flexibility.

In general, the Committee agreed that the attribution approach should be specified for resource use measures (not allowing guidelines) since allowing flexibility would result in different measure results and has implications on comparability. For example, hospital A implements a measure and chooses to attribute all costs for their patients' episodes to the primary care provider (PCP); using the same measure, hospital B chooses to divide the costs of their patients' episodes by attributing costs among all providers that touched the patient. Comparing PCP's at both hospitals in this scenario would be unfair as the rules for how costs are assigned different and disproportionate. Further, some argued that the attribution approach must be described clearly in order to understand the context in which providers are measured and the results are computed, emphasizing the need for the application of the measures to align with the intended use. On the other hand, other members pointed out that given the lack of a gold standard for an attribution method, flexibility in the attribution approach would allow for innovation. The Committee ultimately agreed that resource use measures should move to a more standardized approach of requiring the attribution approach to be submitted as a specification to the measure.

Measure Evaluation Summary

Recommended Measure 14
Measure not recommended 14

Recommended Measure

#2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB) 15

Measure not recommended

#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries 19

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: <u>The measure meets the Importance criteria</u>

(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**

<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:

- 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: <u>The measure meets the Scientific Acceptability criteria</u>

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

(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 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

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

measure results for consumers.

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.

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:

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

#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries
• 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 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. • Significant variation within groups can be masked by group-level reporting; physician-level reporting would eliminate that masking, but presents its own challenges.

#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries
• 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** <u>Rationale:</u> • 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.

Notes

1) Catlin A, et al., National Health Spending in 2006: A Year of Change for Prescription Drugs, Health Affairs, 2008; 27(1):14–29.
2) Banks J, et al., Disease and disadvantage in the United States and in England, JAMA, 2006;295(17):2037–2045.
3) Hoyert DL, et al., Annual summary of vital statistics: 2004, Pediatrics, 2006; 117(1):168–183.
4) Hoyert DL, et al., Annual summary of vital statistics: 2004, Pediatrics, 2006; 117(1):168–183.
5) Weiss JE, Mushinski M, International mortality rates and life expectancy: selected countries, Statistical Bulletin—Metropolitan Life Insurance Company, 1999;80(1):13–21.
6) CMS/Mathematica. (n.d.). *2158: Payment Standardized Medicare Spending per Beneficiary (MSPB).* Retrieved June 27, 2013, from http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=73024
7) Winkelman, Ross, FSA, and Syed Mehmud. *A Comparative Analysis of Claims-Based Tools for Health Risk Assessment*. Rep. Society of Actuaries, 20 Apr. 2007. Web. 28 June 2013. <http://www.soa.org/Files/Research/Projects/risk-assessmentc.pdf>.

Appendix A: Measure Specifications

#2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB) 25

#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries..... 26

	#2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB)
Steward	Centers for Medicare & Medicaid Services (CMS)
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
Construction Logic Description	The MSPB Measure assesses the cost to Medicare of services performed by hospitals and other healthcare providers during an MSPB episode. An MSPB episode is risk adjusted and includes Medicare payments for services provided to a beneficiary with start date falling between 3 days prior to an IPPS hospital admission (index admission) through 30 days post-hospital discharge.
Clinical Framework Description	Objective: The MSPB Measure aims to improve care coordination in the period between 3 days prior to an acute inpatient hospital admission through the period 30 days after discharge. Clinical Topic Area: Inpatient Admissions, all conditions Accounting for Comorbidities: Application of a variant of the CMS-HCC risk adjustment model. The model includes a select number of interaction terms between comorbidities. Measure of Episode Severity: Risk Adjustment model includes indicators for the MS-DRG of the index admission. Concurrency of Clinical Events. The MSPB Episode spans the period 3 days prior to the index hospital admission through 30 days post-discharge. All events that occur during this time period are included in the MSPB episode.
Costing Method	Standardized Pricing
Tested Population	Medicare
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)

	#2158 Payment-Standardized Medicare Spending Per Beneficiary (MSPB)
Attribution Approach	The MSPB episode is attributed to the hospital on the trigger inpatient claim for the index hospital admission that begins an MSPB episode. Specifically, for any period of performance selected, the first set of hospitalizations that can be included in the MSPB Measure are those that begin on the fourth day of the period of performance. This permits sufficient data for the 3-day pre-hospitalization period. Hospitalizations eligible to start an MSPB episode also must end in a discharge 30 days prior to the end of the period of performance to permit the collection of claim information during the post-discharge period. For instance, for the current MSPB figures available on Hospital Compare, the period of performance is May 1, 2011 to December 31, 2011. In this case, hospitalizations that start on May 4 and have a discharge date before December 1 are eligible to be included as index admissions. As discussed in S.9.1., however, due to the uncertainty surrounding attributing episodes to hospitals in cases where the patient was transferred between acute hospitals during the index admission, acute-to-acute transfers during the index admission are not considered index admissions for the purposes of the MSPB Measure. In other words, these cases will not generate new MSPB episodes; neither the hospital which transfers a patient to another short-term acute hospital, nor the receiving short-term acute hospital will have an index admission attributed to them.
Risk Adjustment	Statistical risk model
Stratification	The risk-adjustment model is stratified by major diagnostic category (MDC). MDCs are aggregations of Diagnosis Related Groups (MS-DRG), which CMS uses to classify acute inpatient admissions. The MS-DRG/MDC crosswalk is available for order here: http://solutions9.3m.com/wps/portal/!ut/p/c1/04_SB8K8xLLM9MSSzPy8xBz94NS8-NBg_Qj9KLP4IC8Py1BTI2MD9zAvFwMjYzMzCxNHd2OTACP9ggxHRQBm3gTM/

	#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries
Steward	Centers for Medicare & Medicaid Services (CMS)
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

	#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries
Construction Logic Description	The Payment-Standardized Total Per Capita Cost Measure for Medicare FFS Beneficiaries is formed by first attributing beneficiaries to medical group practices. Then, unadjusted per capita costs are calculated as the sum of all Medicare Part A and Part B costs for all beneficiaries attributed to a medical group practice, divided by the number of attributed beneficiaries. All unadjusted costs are then payment-standardized and risk adjusted to accommodate differences in costs between peers that result from circumstances beyond physicians' control. Risk-adjusted costs are computed as the ratio of a medical group practice's payment-standardized (but not risk-adjusted) per capita costs to its expected per capita costs, as determined by the risk adjustment algorithm. Finally, to express the risk-adjusted cost in dollars and for ease of interpretation, the ratio is multiplied by the mean cost of all beneficiaries attributed to all practices.
Clinical Framework Description	This is an annual payment-standardized per capita cost measure for medical group practices that applies to all clinical topic areas. Comorbidities and clinical hierarchies are accounted for during the risk-adjustment process.
Costing Method	Standardized Pricing
Tested Population	Medicare; Medicaid
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; Hospice; Home health; skilled nursing facility; Anesthesia; Ambulance services; Chemotherapy; Drugs administered in an ambulatory setting or used with DME (covered by Medicare Part B); Orthotics, chiropractic, enteral and parenteral nutrition; some vision services; some hearing and speech services; immunizations
Attribution Approach	Beneficiaries are attributed to medical group practices that provided the plurality of primary care services (PCS). Only beneficiaries that received PCS from at least one physician during the measurement period are eligible for assignment.
Risk Adjustment	Statistical risk model
Stratification	This measure uses risk-adjusted costs for comparison purposes and further stratification is not done.

Appendix B: Project Steering Committee and NQF Staff

STEERING COMMITTEE

Eugene Nelson, DSc, MPH (Co-Chair)
Dartmouth Institute For Health Policy and Clinical Practice
Norwich, VT

David Penson, MD, MPH (Co-Chair)
Vanderbilt University
Nashville, TN

Lawrence Becker
Xerox Corporation
Rochester, NY

Mary Ann Clark, MHA
Intralign
Washington, DC

Cheryl Damberg, PhD
RAND Corporation
Santa Monica, CA

Jennifer Eames-Huff, MPH
Pacific Business Group on Health
San Francisco, CA

Nancy Garrett, PhD
Hennepin County Medical Center
St. Paul, MN

Andrea Gelzer, MD, MS, FACP
AmeriHealth Mercy Family of Companies
Philadelphia, PA

David Gifford, MD, MPH
American Health Care Association
Washington, DC

Stanley Hochberg, MD
Boston Medical Center
Boston, MA

Lisa Latts, MD, MSPH, MBA, FACP
LML Health Solutions, LLC
Denver, CO

Matthew McHugh, PhD, JD, MPH, RN, CRNP, FAAN
University of Pennsylvania
Philadelphia, PA

Martin Marciniak, MPP, PhD
GlaxoSmithKline
Research Triangle Park, NC

James Naessens, ScD, MPH
Mayo Clinic
Rochester, MN

Jack Needleman, PhD
UCLA Fielding School of Public Health
Los Angeles, CA

Carolyn Pare
Minnesota Health Action Group
Bloomington, MN

David Redfearn, PhD
Wellpoint
Las Vegas, NV

Andrew Ryan, PhD
Weill Cornell Medical College
New York, NY

Joseph Stephansky, PhD
Michigan Health & Hospital Association
East Tawas, MI

Thomas Tsang, MD, FACP
Merck
Boston, MA

Lina Walker, PhD
AARP – Public Policy Institute
Washington, DC

William Weintraub, MD, FACC
Christiana Care Health System
Newark, DE

Daniel Wolfson, MHSA
ABIM Foundation
Philadelphia, PA

Herbert Wong, PhD
Agency for Healthcare Research and Quality
Rockville, MD

Dolores Yanagihara, MPH
Integrated Healthcare Association
Oakland, CA

NQF STAFF

Helen Burstin, MD, MPH
Senior Vice President
Performance Measures

Taroon Amin, MA, MPH
Senior Director
Performance Measures

Ashlie Wilbon, RN, MPH
Senior Project Manager
Performance Measures

Lindsey Tighe, MS
Project Manager
Performance Measures

Evan M. Williamson, MPH, MS
Project Analyst
Performance Measures

Appendix C: Measures Endorsed in Cost and Resource Use Since April 2012

NQF Number	Title	Steward
1557	Relative Resource Use for People with Diabetes	National Committee for Quality Assurance (NCQA)
1558	Relative Resource Use for People with Cardiovascular Conditions	National Committee for Quality Assurance (NCQA)
1560	Relative Resource Use (RRU) for People with Asthma	National Committee for Quality Assurance (NCQA)
1561	Relative Resource Use for People with Chronic Obstructive Pulmonary Disease (COPD)	National Committee for Quality Assurance (NCQA)
1598	Total Resource Use Population-based PMPM Index	HealthPartners
1604	Total Cost of Care Population-based PMPM Index	HealthPartners
1609	ETG based Hip/Knee Replacement Cost of Care	Ingenix/OptumInsight
1611	ETG based Pneuomonia Cost of Care	Ingenix/OptumInsight

Appendix D: Related and Competing Measures

Comparison of NQF #1598 and NQF #2165

	#1598 Total Resource Use Population-based PMPM Index	#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries
Steward	HealthPartners	Centers for Medicare & Medicaid Services (CMS)
Description	The Resource Use Index (RUI) is a risk adjusted measure of the frequency and intensity of services utilized to manage a provider group's patients. Resource use includes all resources associated with treating members including professional, facility inpatient and outpatient, pharmacy, lab, radiology, ancillary and behavioral health services.	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)	Per capita (population- or patient-based)
Data Source	Administrative claims, Other: Users administrative claims data base, Risk-adjustment Tool, Johns Hopkins ACG System Version 9.0, Standardized costing code table, Total Care Relative Resource Values (TCRRV) specification provided	Administrative Claims
Level of Analysis	Clinician: Group/Practice; Population: Community	Clinician : Group/Practice

	#1598 Total Resource Use Population-based PMPM Index	#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries
Construction Logic Description	The measure examines total resource use of a commercial population between for a given measurement year (e.g. January 1 and December 31), for all members eligible for the measure	The Payment-Standardized Total Per Capita Cost Measure for Medicare FFS Beneficiaries is formed by first attributing beneficiaries to medical group practices. Then, unadjusted per capita costs are calculated as the sum of all Medicare Part A and Part B costs for all beneficiaries attributed to a medical group practice, divided by the number of attributed beneficiaries. All unadjusted costs are then payment-standardized and risk adjusted to accommodate differences in costs between peers that result from circumstances beyond physicians' control. Risk-adjusted costs are computed as the ratio of a medical group practice's payment-standardized (but not risk-adjusted) per capita costs to its expected per capita costs, as determined by the risk adjustment algorithm. Finally, to express the risk-adjusted cost in dollars and for ease of interpretation, the ratio is multiplied by the mean cost of all beneficiaries attributed to all practices.
Clinical Framework Description	Not applicable. This is a population-based measure that applies to all service categories, care settings and conditions.	This is an annual payment-standardized per capita cost measure for medical group practices that applies to all clinical topic areas. Comorbidities and clinical hierarchies are accounted for during the risk-adjustment process.

	#1598 Total Resource Use Population-based PMPM Index	#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries
Costing Method	**Description:** The Total Care Relative Resource Values (TCRRVs) are a grand linear scale of relative values designed to evaluate resource use across all types of medical services, procedures and places of service. The values are independent of price and can be used to evaluate providers, hospitals, physicians and health plans against their peers on their efficiency of resource use in treating like conditions. **General Overview of Application:** The TCRRVs are applied at the procedure level for each component of care with the exception of inpatient, which is applied at the full admission level. There is a TCRRV lookup table for each component of care where each claim's procedure is matched with the corresponding value. The TCRRV weights that are applied to the claim is tested for accuracy and a total TCRRV is calculated. The final step is to calibrate the total TCRRVs to the paid ratio between components of care using the paid adjustment factor. www.healthpartners.com/files/56500.pdf OR www.healthpartners.com/tcoc.	Standardized Pricing
Tested Population	Commercial	Medicare; Medicaid

	#1598 Total Resource Use Population-based PMPM Index	#2165 Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries
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: Pharmacy; Ambulatory services: Evaluation and management; Ambulatory services: Procedures and surgeries; Ambulatory services: Imaging and diagnostic; Ambulatory services: Lab services; Ambulatory services: Labor (hours, FTE, etc.); Durable Medical Equipment (DME)	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; Hospice; Home health; skilled nursing facility; Anesthesia; Ambulance services; Chemotherapy; Drugs administered in an ambulatory setting or used with DME (covered by Medicare Part B); Orthotics, chiropractic, enteral and parenteral nutrition; some vision services; some hearing and speech services; immunizations
Attribution Approach	**Guidelines:** To determine which members to include in the Total Resource Use measure, there are several options available depending upon your business purpose and unit of measure. If the unit of measure is an entire health plan or employer group, all members will be included in the Total Resource Use measure. If the unit of measure is a provider and members are required to select a primary care provider, we recommend using the member selected provider. When the member is not required to select a primary care provider, we recommend the use of an attribution algorithm to identify the member's primary care provider. The measure was tested using this methodology.	Beneficiaries are attributed to medical group practices that provided the plurality of primary care services (PCS). Only beneficiaries that received PCS from at least one physician during the measurement period are eligible for assignment.
Risk Adjustment	For Total Resource Use measurement, risk adjustment is performed using Adjusted Clinical Groups (ACG) developed by Johns Hopkins University.	Statistical risk model
Stratification	This is a population-based measure that is fully inclusive.	This measure uses risk-adjusted costs for comparison purposes and further stratification is not done.

ACUMEN
500 Airport Blvd., Suite 365
Burlingame, CA 94010
650-558-8882
Acumenllc.com

MEMORANDUM

TO:	Cost and Resource Use Steering Committee National Quality Forum
FROM:	Thomas MaCurdy, Sajid Zaidi, David Pham, Elen Shrestha, Leah Rosenbaum, and Lynn Redington Acumen, LLC
CC:	Kimberly Spalding Bush, Craig Caplan, and Michael Wroblewski Centers for Medicare & Medicaid Services
DATE:	June 27, 2013
REFERENCE:	Responses to Cost and Resource Use Steering Committee Concerns Medicare Spending per Beneficiary (MSPB) Measure (2158)

CMS submitted the Medicare Spending per Beneficiary (MSPB) measure to the National Quality Forum (NQF) for endorsement on January 31, 2013. During the May 8-9, 2013 NQF Cost and Resource Use Steering Committee Meeting, the Committee voted to recommend endorsement of the MSPB measure and also identified areas of concern in the MSPB measure's NQF Measure Submission Form. We thank the Committee for their thoughtful consideration of this measure and for the additional research questions they posed. Their suggestions have facilitated a more robust analysis of the MSPB measure. We also thank the National Quality Forum for the opportunity to submit these additional analyses and findings, as well as some clarifications to our initial submission. We believe that you will find the results in this memorandum and appendix support that the MSPB measure is highly reliable and valid for the measurement of Medicare spending surrounding hospitalizations. Accordingly, we do not intend to change the measure's specifications at this time, but will continue analyses for potential future refinements.

The Executive Summary presents a brief description of Acumen's analyses, responses, and clarifications related to validity and reliability concerns expressed by the Committee. Afterwards, in the Detailed Analyses section, Committee concerns and comments discussed in the Executive Summary are addressed in additional detail, and each section has a short summary. Throughout this memo, all references to cost refer to price-standardized Medicare payments. "Observed" cost refers to non-risk-adjusted, price-standardized Medicare payment, while "risk-

adjusted" cost refers to risk-adjusted, price-standardized Medicare payment. This memo uses data from the May 2011 – December 2011 period of performance except when otherwise noted.

EXECUTIVE SUMMARY

Most of the Committee's concerns were related to scientific acceptability (validity and reliability). However, the Committee was also concerned that the language in the submission form may give the impression that the MSPB measure is a care coordination measure. Acumen and CMS agree with the Committee that the MSPB measure is not a care coordination measure, but is rather a Medicare payment measure. Acumen wishes to clarify that the mention of care coordination in the NQF Measure Submission Form was intended as an example of one area that hospitals could improve in order to reduce Medicare spending during the episode and thereby improve performance scores on the MSPB measure. Acumen also wishes to clarify that the MSPB measure is a measure of costs to Medicare or Medicare payment, not a measure of costs to providers. Below is a summary of the Committee's concerns, along with the associated analyses and findings.

Measure Validity

With regard to testing the validity of the MSPB measure, the Committee asked for analyses to better understand how the MSPB measure correlates with other measures and how the measure varies among selected hospital and patient strata. The Committee also asked for additional analyses of exclusions, specifically exclusions of acute-to-acute (hospital) transfers, outliers, and death episodes, and asked for analyses that examined selected aspects of the MSPB risk adjustment methodology. Finally, the Committee questioned the measure's construction using Part A and Part B data, but not Part D.

1. **Correlations with Other Cost Measures:** The Committee suggested that analysis of correlation between the MSPB measure and other cost measurement data would support the MSPB measure's validity.

   Analysis: Correlation with an overall service utilization measure.
   Result: There is a positive, statistically significant correlation with the MSPB measure of 0.22.

   Analysis: Correlation with Hospital Referral Region (HRR)-level aggregate risk-adjusted, annual per capita spending for all Medicare beneficiaries enrolled in Medicare fee-for-service (originally calculated for the Institute of Medicine's geographic variation in Medicare Spending, Utilization, and Quality project).

Result: There is a positive, statistically significant correlation of 0.55, meaning that hospitals with more expensive MSPB episodes are generally located in HRRs with higher risk-adjusted annual per capita Medicare spending.

Analysis: Correlation with HRR-level aggregate price-standardized, risk-adjusted, annual per capita spending for all privately insured under-65 members in the Marketscan database, a large commercial claims database (originally calculated by Harvard University researchers for the Institute of Medicine's geographic variation in Medicare Spending, Utilization, and Quality project).
Result: There is a positive, statistically significant correlation of 0.37 with the aggregate per capita spending, meaning that hospitals with more expensive Medicare MSPB episodes are generally located in HRRs with higher price-standardized, risk-adjusted annual per capita spending for commercially insured members.

Analysis: Correlation with subsets of the Marketscan database members, Acute Myocardial Infarction (AMI), and Stroke cohorts.
Result: For the AMI cohort, there is a positive, statistically significant correlation of 0.14, and for the stroke cohort, there is a positive, statistically significant correlation of 0.28, meaning that hospitals with more expensive Medicare MSPB episodes are generally located in HRRs with higher price-standardized, risk-adjusted spending for commercially insured members with AMI and stroke hospitalizations.

Conclusion: The results of these analyses indicate that the Medicare MSPB measure is correlated with measures of cost for both the Medicare population aggregated by HRR and for a completely separate group of patients, the under-65 commercially insured population. This gives confidence that the MSPB measure is measuring underlying patterns of utilization, further lending support to the validity of the MSPB measure.

Future Analysis: Correlation with an equivalent to the MSPB measure using Medicaid claims data for beneficiaries who were not dually eligible for both Medicare and Medicaid (not "dual eligibles"). Acumen will complete this analysis and submit it in a memo by August 7th.

2. **Stratifications by Characteristics:** The Committee requested additional information on MSPB measure rate by hospital and patient characteristics.

   Analysis: Stratification by hospital characteristics.
   Result: Larger hospitals, urban hospitals, hospitals with more Medicare patients, teaching hospitals, and hospitals in the south and northeast have more expensive MSPB episodes.

Analysis: Stratification by different patient subpopulations.
Result: Women have higher risk-adjusted spending than men; black beneficiaries have higher risk-adjusted spending than white beneficiaries, and dual eligible beneficiaries (a proxy for socioeconomic status) have higher risk-adjusted spending than non-dual eligible beneficiaries.

Conclusion: The results of these analyses by characteristics are consistent with findings in the literature, supporting the validity of the MSPB measure.

3. **Exclusions:** The Committee expressed concern that the MSPB measure's validity could potentially be adversely affected because a portion of Medicare spending data is lost through exclusion of: acute-to-acute transfers, outlier episodes, episodes in which the beneficiary dies, and Medicaid payment data.

   Analysis: Impact of including acute-to-acute transfer episodes and attributing them to both the receiving and transferring hospitals.
   Result: MSPB measure results that include these transfers are highly correlated with MSPB measure results that do not (0.97 and 0.99, depending on the weighting method).

   Analysis: Impact on the MSPB measure if outliers were fully included rather than excluded.
   Result: Comparing inclusion to exclusion of outliers, Acumen found the MSPB measure results are highly correlated (0.95).

   Analysis: Impact of including death episodes and including a death indicator in the risk adjustment model (submitted with initial Measure Submission Form, and not repeated in the "Detailed Analysis" section below).
   Result: When including a death risk-adjustment variable, average expected cost of a death episode falls to $22,706, and average expected cost of a non-death episode is $19,495. 19.7 percent of hospitals experience a change in their MSPB measure of more than 3 percentage points when adding death episodes in this way.

   Conclusion: The results of these analyses indicate that exclusion of transfers and outlier episodes have very little effect on the ranking of hospitals in the MSPB measure and therefore do not adversely affect its validity.

   With regard to the exclusion of episodes in which the beneficiary dies, this exclusion was finalized through notice and comment rulemaking, based on the fact that these are incomplete episodes where significant data could be missing when death occurs early in the episode. Further, unusually high expenses for end of life care may exist when death

occurs at the end of the episode. CMS will consider including episodes in which the beneficiary dies in future updates to the MSPB measure.

With regard to the exclusion of Medicaid data, Acumen would like to assure the Committee that no Medicare payment data would be missing for beneficiaries who are also covered by Medicaid. This is because Medicare is always a primary payer to Medicaid for Medicare covered services, so Medicare payments for these services would always appear in Medicare claims files. The MSPB measure includes only Medicare spending, consistent with the requirement in 1886(o)(2)(B)(ii), as added by section 3001 of the Affordable Care Act, that the Hospital Value-Based Purchasing include measures of "Medicare spending per beneficiary."

4. **Risk Adjustment Methodologies:** Committee members requested clarification as to whether a 90-day look-back period for risk adjustment was sufficient and whether Present on Admission (POA) diagnoses from the index admission should be included in the risk adjustment model. The Committee also questioned the decile analysis presented in the initial submission. One Committee member also asked if Acumen could test using the natural log of spending, instead of the level.

   Analysis: The MSPB measure was recalculated, including all diagnoses that were present on admission in the risk adjustment model.
   Result: Including POA diagnoses did not materially change the R^2 of the regression (0.45 to 0.46).

   Analysis: Revised approach to analysis of model calibration and decile plot analyses.
   Result: The risk adjustment model performs well at discriminating between high cost and low cost episodes and at predicting cost throughout the distribution.

   Analysis: Natural log risk adjustment model.
   Result: Using the natural log worsens the fit of the model. The R^2 is 0.41 when observed episode cost is the dependent variable in the regression, while the R^2 is 0.39 when the natural log of observed episode cost is the dependent variable in the regression.

   Analysis: Coefficient of Determination (R^2) with a 365-day look-back period (submitted with initial Measure Submission Form).
   Result: Switching from a 90-day look-back period to a 365-day look-back did not materially change the R^2 of the regression (0.4621 to 0.4601).

   Conclusion: This result, along with the analyses provided in the NQF Measure Submission Form, support that including POA diagnoses or a longer look-back period have very little impact on risk adjustment model performance or final MSPB measure

scores and that the current risk adjustment model performs well in predicting MSPB cost. Further, the new decile analysis supports the validity of the risk adjustment methodology.

5. **Source of Cost Variation:** The Committee asked what proportion of the variation in risk-adjusted MSPB episode cost is due to post-discharge costs and what portion is due to index admission costs.

   Analysis: The variance in total risk-adjusted MSPB episode cost was decomposed into variance in post-discharge cost and variance in index admission costs.

   Result: Variance of risk-adjusted post-discharge cost accounts for approximately 80 percent of total risk-adjusted cost variance.

   Conclusion: After risk-adjustment, most of the remaining cost variation is due to cost variation in the post-discharge window. It is important to note that the risk adjustment does adjust for each beneficiary's predicted level of post-discharge spending based on prior health history and the MS-DRG; the variance in post-discharge cost that remains is unaccounted for by the beneficiary's risk.

6. **Part D Data:** The Committee expressed concern that the MSPB measure is constructed of Part A and Part B data, but not Part D. While we appreciate that Part D data represent a significant Medicare expenditure, we are unable to include Part D data until a standardization approach can be fully vetted through stakeholders, similar to the Part A and B standardization methodology. We intend to further analyze the inclusion of Medicare Part D data for potential future refinement and resubmission of this measure.

Measure Reliability

Committee members expressed concerns about MSPB measure reliability analyses, specifically Acumen's test/retest analysis and utilization of an 8-month period of performance for MSPB measure calculations vs. a 12-month period of performance.

1. **Test/Retest Analysis:** The Committee expressed concern with the "test/retest analysis," in which beneficiaries are randomly split into two non-overlapping samples, and MSPB measures are statistically compared. By comparing the correlation of a hospital's MSPB measure calculated using the two mutually exclusive samples, one can identify the precision of a hospital's score across multiple random samples. Specifically, the Committee was concerned that out of the hospitals in the top quintile in one sample, 30 percent were not in the top quintile in the other sample.

Clarification: Acumen wishes to clarify the findings. The result that 70 percent of hospitals in the top quintile in one sample remain in the top quintile in the other sample is evidence of a highly stable measure, and is a high figure by the standards of quintile stability analyses (for comparison, only 20 percent of hospitals are expected to remain in the same quintile by random chance). In addition, 90 percent of the hospitals in the top quintile in one sample remain in the top two quintiles in the other sample. Finally, the strong, statistically significant rank correlation of 0.84 between the two samples also indicates a stable, precise measure.

Conclusion: The test/retest findings indicate a stable, reliable measure across multiple random, non-overlapping samples. This conclusion is supported by Carlos Alzola, the NQF's statistical consultant, who said during the NQF Cost and Resource Use Steering Committee Meeting that the Spearman rank correlation was more than sufficient and satisfied him with respect to reliability.

Period of Performance Analysis: The Committee questioned whether an 8-month period of performance was similar to a 12-month period of performance for the MSPB measure.

Analysis: Used both an 8-month period of performance and a 12-month period of performance to calculate the MSPB measure.
Result: The resulting sets of scores are highly correlated (0.97).

Conclusion: An 8-month period of performance is sufficient for the MSPB measure as it is highly correlated with a 12-month period.

DETAILED ANALYSES

Measure Validity

1. Correlation with Other Cost Measures

Summary: Acumen calculated an overall service utilization measure and found that it has a positive, statistically significant correlation with the MSPB measure of 0.22. Acumen also compared the MSPB measure with HRR-level aggregate risk-adjusted annual per capita spending for all Medicare beneficiaries enrolled in Medicare fee-for-service (originally calculated for the Institute of Medicine's geographic variation in Medicare Spending, Utilization, and Quality project) and found a positive, statistically significant correlation of 0.55. Acumen also compared the MSPB measure with HRR-level aggregate price-standardized, risk-adjusted annual per capita spending for the under-65, commercially insured population and found a positive, statistically significant correlation of 0.37. The correlations with AMI and Stroke post-hospitalization spending for the same commercially insured population are 0.14 and 0.28, respectively, and are both statistically significant. These numbers all show that the MSPB measure is correlated with other measures of price-standardized, risk-adjusted cost, supporting the validity of the MSPB measure.

Acumen originally submitted hospital-level correlations of the MSPB measure with the three CMS 30-day readmission measures for heart failure, pneumonia, and AMI. These were positive and statistically significant, but low in magnitude. This result was a source of concern for the Committee; however, the low correlation may be explained by the fact that the MSPB measure is an all-cause measure that includes all spending, while the readmission measures are for only three conditions and only measured inpatient readmission.

To supplement this analysis and address the Committee's concerns, Acumen subsequently constructed utilization measures for various categories of medical services. Acumen found a statistically significant and strong positive correlation of 0.6 with both professional evaluation and management (E&M) services, post-acute services (including inpatient hospital (IP), and skilled nursing facility (SNF)), which together account for a majority of medical spending. Correlation with utilization in other service categories (e.g. Procedure Services, Other Hospital services, Emergency Services, and Ancillary Services) were smaller, but were all still positive and statistically significant. During the Committee meeting, NQF's consulting statistician Carlos Alzola stated that the correlation of 0.6 is acceptable and increased his level of confidence in the validity of the measure.

Since then, Acumen has calculated a combined utilization measure which combines all the categories listed above together and serves as a proxy for overall utilization. Acumen found that the MSPB measure exhibits a positive, statistically significant Pearson correlation of 0.22 with the combined utilization of services categories. This positive correlation indicates that, as would be expected, hospitals with more expensive MSPB episodes generally have higher combined utilization of services. This finding lends further support to the validity of the MSPB measure. Table A in the appendix of this memorandum presents this result, as well as the results from Acumen's previous analysis.

Under the direction of the Institute of Medicine, Acumen previously examined geographic variation in the volume and intensity of annual per capita health care services and spending for both Medicare and Medicaid beneficiaries as part of the Medicare Spending, Utilization, and Quality project.[1] Comparing the MSPB measure using a period of performance from May 2010 to February 2011 with 2009 HRR-level aggregate risk-adjusted annual per capita spending for all Medicare beneficiaries enrolled in Medicare fee-for-service, Acumen found a positive, statistically significant correlation of 0.55. Comparing the MSPB measure using a period of performance from May 2011 to December 2011 with the same HRR-level aggregate risk-adjusted annual per capita spending gives a similar positive, statistically significant correlation of 0.55. This positive correlation indicates that, as would be expected, hospitals with more expensive MSPB episodes are generally located in HRRs with higher risk-adjusted annual per capita spending. This finding lends further support to the validity of the MSPB measure.

Acumen also correlated the MSPB measure with the results of the IOM analysis of privately insured, under-65 members from the large Marketscan database.[2] This analysis, conducted by Harvard University researchers on behalf of the IOM, price-standardized and risk-adjusted the private Marketscan claims data. Comparing the Medicare MSPB measure using a period of performance from May 2010 to February 2011 with 2009 HRR-level aggregate price-standardized, risk-adjusted annual per capita spending for Marketscan members, Acumen found a positive, statistically significant correlation of 0.37. Comparing the same Medicare MSPB measure with price-standardized, risk-adjusted post-hospitalization costs for Marketscan members who had an AMI or Stroke, Acumen found positive, statistically significant correlations of 0.14 and 0.28, respectively. Because the Marketscan analysis used entirely different data sources, payment systems, and population than Medicare, these correlations serve

[1] MaCurdy, Thomas, et al. "Geographic Variation in Spending, Utilization and Quality: Medicare and Medicaid Beneficiaries." Burlingame, CA: Acumen, LLC. May 2013. http://iom.edu/Reports/2013/~/media/Files/Report%20Files/2013/Geographic-Variation/Sub-Contractor/Acumen-Medicare-Medicaid.pdf

[2] McKellar, Michael, et al. "Geographic Variation in Health Care Spending, Utilization, and Quality Among the Privately Insured." Boston, MA: Harvard Medical School Department of Health Care Policy. August 2012 http://www.iom.edu/Reports/2013/~/media/Files/Report%20Files/2013/Geographic-Variation/Sub-Contractor/Harvard-University.pdf

as evidence that the MSPB measure is consistently capturing underlying patterns of cost, supporting its validity as a cost measure.

2. Stratifications by Characteristics

Summary: The Committee asked to see more stratifications of the MSPB measure in order to compare the results against findings in the literature. Our findings that larger hospitals, urban hospitals, hospitals with more Medicare patients, teaching hospitals, and hospitals in the south and northeast are more expensive are consistent with the literature. In addition, Committee members asked to see stratifications of the MSPB measure by gender, race, and socioeconomic status. We find that women have higher risk-adjusted spending than men, that black beneficiaries have higher risk-adjusted spending than white beneficiaries, and that dual eligible beneficiaries (a proxy for socioeconomic status) have higher risk-adjusted spending than non-dual eligible beneficiaries. These findings are consistent with findings in the literature, supporting the validity of the MSPB measure. Acumen also originally found that at the episode level, dual eligible beneficiaries cost more than beneficiaries who are eligible for Medicare only, for both non-risk-adjusted and risk-adjusted cost.[3] However, this relationship was not evident at the hospital level, and the Committee questioned this result, asking whether it was a validity concern. Since then, further analysis suggests that it is likely a result of confounding factors at the hospital level. When these confounding factors are controlled for, hospitals with more dual eligible beneficiaries do have higher spending. A simple hospital-level regression shows that hospitals with more dual eligible beneficiaries do indeed have more expensive MSPB episodes (with a statistically and practically significant positive coefficient) after controlling for teaching status, hospital size, and urban/rural location.

In response to the Committee request to calculate more stratifications of the MSPB measure, we made the following observations with regard to hospital attributes:

- Hospital Size (number of beds): There is a statistically significant correlation of 0.16 between the MSPB measure and the number of beds in a hospital (Table 1).
- Percent of Total Inpatient Days that are for Medicare Patients: There is a statistically significant correlation of 0.04 between the MSPB measure and the percentage of inpatient days for Medicare patients (Table 1).
- Urban vs. Rural Hospitals: On average, urban hospitals have a higher MSPB measure than rural hospitals. Average observed spending per episode is also

[3] As stated in the submission form, the risk adjustment methodology for the MSPB measure does not adjust for dual eligibility.

higher for urban hospitals than for rural hospitals, with a spending difference of approximately $3,000 (Appendix Table B).

- Region: West North Central (includes IA, KS, MN, MO, NE, ND, and SD) has the lowest average MSPB measure (0.93) and the lowest average observed spending per episode ($17,807). New England (includes CT, MA, ME, NH, RI and VT) and West South Central (includes AR, LA, OK, and TX) both have the highest average MSPB measure (1.01) (Appendix Tables B and C).
- Teaching Hospitals: On average, teaching hospitals have slightly higher MSPB measures than non-teaching hospitals. Average observed spending per episode is also higher for teaching hospitals than for non-teaching hospitals.

These findings confirm what is found in the literature, lending further support to the validity of the MSPB measure. For example, the Dartmouth Atlas of Health Care shows that the Great Plains states generally have the lowest Medicare utilization, while southern and northeastern states have the highest. Urban areas also have higher Medicare utilization.[4] Larger hospitals and academic centers have also been shown in the literature to have higher Medicare spending.[5] For additional information regarding these analyses, please refer to the workbook titled "NQF_MSPB_Correlation_Analysis_09JUN2013" attached with this memorandum.

Table 1: MSPB Measure Correlations by Hospital Size and Percent of Inpatient Days for Medicare Patients

	MSPB Measure	Observed Cost per Episode
	Correlation	**Correlation**
Hospital Size (# of Beds)	0.16	0.41
% of IP Days for Medicare Patients	0.04	-0.17

Some members of the Committee expressed concerns that the MSPB measure does not adjust for sex, race, or socioeconomic factors in the risk adjustment methodology. As noted in the NQF Measure Submission Form, this decision is consistent with NQF's position on not adjusting for demographic (sex or race) or socioeconomic factors when there is a potential disparity in care. In order to examine the effect of these factors on MSPB amounts, as suggested by the Steering Committee, Acumen stratified MSPB amounts by sex and race (socioeconomic

[4] Skinner, Jonathan et al. "A New Series of Medical Expenditure Measures by Hospital Referral Region: 2003-2008". The Dartmouth Institute for Health Policy and Clinical Practice. June 21, 2011. http://www.dartmouthatlas.org/downloads/reports/PA_Spending_Report_0611.pdf

[5] Romley, John et al. "Spending and Mortality in US Acute Care Hospitals." Am J Manag Care. 2013;19(2):e46-e54

status is addressed below in the dual eligibles section).[6] The results show disparities along these factors, consistent with current research, lending further support to the measure's validity.

Acumen found that men have higher observed spending, but lower risk-adjusted spending than women (Table 2). This is consistent with the literature that indicates that women generally have higher health care spending.[7] It is important to note that the risk adjustment controls for the MS-DRG of the index admission, which indicates the reason for hospitalization. Thus, gender differences in the incidence of disease (such as breast cancer or prostate cancer) would not result in MSPB measure differences, since these are risk adjusted out.

When examining racial differences, Acumen found that Asians have the highest observed spending, while Native Americans have the lowest observed spending. On the other hand, black beneficiaries have the highest risk-adjusted spending, while Native American beneficiaries also have the lowest risk-adjusted spending (Table 3).

Table 2: MSPB Amount Breakdown by Sex

Gender	%	Observed	Risk-Adjusted
Female	58%	$18,263	$18,524
Male	42%	$18,488	$18,140

Table 3: MSPB Amount Breakdown by Race

Race	%	Observed	Risk-Adjusted
Asian	1%	$18,616	$17,964
Black	12%	$18,592	$18,499
Hispanic	2%	$17,961	$18,044
Native American	1%	$16,635	$16,984
Other	1%	$18,329	$17,870
White	82%	$18,343	$18,368

These findings are consistent with the literature on racial disparities in health care spending. As stated in the NQF Measure Submission Form, end-of-life care for black and Hispanic beneficiaries is substantially different than the end-of-life hospital services that white Medicare beneficiaries receive. Much of the variation is due to differences in utilization levels among hospitalized patients. Black and Hispanic patients are significantly more likely to be admitted to the ICU than are white patients, and minority patients also receive significantly more intensive procedures, such as resuscitation and cardiac conversion, mechanical ventilation, and

[6] The MSPB Measure is calculated as the ratio of the MSPB amount for a hospital divided by the median MSPB amount across all hospitals where the MSPB amount is defined as the average price-standardized, risk-adjusted spending across all of the hospital's eligible episodes.

[7] Owens, GM. "Gender differences in health care expenditures, resource utilization, and quality of care." J. Manag. Care Pharm. 2008 Apr;14(3 Suppl):2-6.

gastrostomy for artificial nutrition.[8] Further, there also exists significant variation in the inpatient procedures received by patients of different races. White patients, for example, get almost three times as many carotid endarterectomies as black patients, and 30 percent more angiograms. On the other hand, black patients have higher rates of admission to the ICU in their last six months of life. On average, black enrollees have more money spent on them, particularly near the end of life, but receive fewer highly effective interventions.[9] In addition, a number of studies have shown that the quality of post-acute care varies across patient socioeconomic status. For example, an analysis of 30-day readmission rates revealed that among the Medicare population, black beneficiaries were more likely to be readmitted after hospitalization for AMI, congestive heart failure (CHF), and pneumonia, a gap that was related to both race and to the site where care was received. Specifically, black beneficiaries had higher readmission rates than white beneficiaries across all three conditions, and patients from minority-serving hospitals had higher readmission rates than non-minority-serving hospitals.[10] Whereas one quarter of Medicare beneficiaries with incomes less than $20,000 per year used inpatient services in a given year, only 17 percent of patients earning over $30,000 per year used inpatient services. Beneficiaries with incomes below $20,000 are also twice as likely to use home health services as Medicare beneficiaries earning more than $30,000.[11] This literature confirms the validity of the differences in MSPB measure by race, gender, and socioeconomic status.

Table 4 shows that dual eligibles are more expensive at the episode level on both non-risk-adjusted cost and risk-adjusted cost (the differences are both statistically and practically significant). However, at the hospital level, there is actually a negative relationship between the percentage of beneficiaries who are dual eligible and the non-risk-adjusted cost (Table 5), while there is a small positive but not statistically significant relationship with risk-adjusted cost (Table 6).

Table 4: Episode-Level Costs

	Dual Eligible	**Non-Dual Eligible**
Mean Observed Cost	$18,680	$18,206
Risk-Adjusted Cost	$18,802	$18,150

[8] Hanchate, Amresh, et al. "Racial and Ethnic Differences in End-of-Life Costs: Why do Minorities Cost More than Whites?" Archives of Internal Medicine. 2009; 169(5):493-504.
[9] Baicker, Katherine, et al. "Who You Are and Where You Live: How Race and Geography Affect the Treatment of Medicare Beneficiaries." Health Affairs, October 2004.
[10] Joynt, Karen, et al. "Thirty-Day Readmission Rates for Medicare Beneficiaries by Race and Site of Care." JAMA. February 2011; 305(7): 675-681.
[11] [1] Kaiser Family Foundation. "Medicare Chartbook" Fourth Edition, 2010. http://www.kff.org/medicare/upload/8103.pdf

Table 5: Hospital-Level Regression of Mean Observed Cost on Percent of Dual Eligibles

	Intercept	Dual Eligible %
Coefficient	18713.19	-4612.72
P-value	0.0	0.0

Table 6: Hospital-Level Regression of Risk-Adjusted Cost on Percent of Dual Eligibles

	Intercept	Dual Eligible %
Coefficient	17921.70	213.64
P-value	0.0	0.22

The Committee was concerned about these hospital level results, as they expected hospitals with more dual eligible to be more expensive. However, after controlling for a few confounding factors at the hospital level, the percent of beneficiaries who are dual eligible does have a positive and statistically significant relationship with risk-adjusted cost (Table 7). This indicates that, at the hospital level, these other factors were themselves correlated with the percent of beneficiaries who were dual eligible, and only after controlling for them do we see the expected positive relationship. This analysis shows that, as expected, dual eligible beneficiaries are more expensive at both the episode level and at the hospital level (when adjusted for confounding factors), supporting the validity of the MSPB measure. For additional information regarding this analysis, please refer to the workbook titled "NQF_Dual_Eligible_Cost_Analysis_13JUN2013" attached with this memorandum.

Table 7: Hospital-Level Regression of Risk-Adjusted Cost on Multiple Factors

	Intercept	Dual Eligible %	Teaching Status	Urban Status	# of Beds
Coefficient	17009.50	592.39	-295.03	861.64	1.31
P-value	0.0	0.05	0.02	0.0	0.0

The differences in MSPB spending along gender and racial lines, as well as along socioeconomic status, are consistent with findings in the literature and further support the validity of the measure.

3. Exclusions

Summary: Some Committee members expressed concern with the exclusion of acute-to-acute transfer cases from initiating MSPB episodes. Acumen previously submitted analyses showing that transfer episodes are much more expensive than other episodes, and that small rural facilities are more likely to transfer patients than are urban facilities. Acumen also previously analyzed the effects of attributing episodes to the transferring facility and of attributing to the receiving facility, and found that the former

disproportionally disadvantaged rural facilities, while the latter disadvantaged large urban facilities (although the effect was of lower magnitude, due to a larger patient population). Acumen has subsequently analyzed attributing transfer episodes to both the transferring and receiving hospitals and found that the results are highly correlated (0.97 and 0.99) with the current MSPB measure excluding transfers. This finding indicates that the MSPB measure is not very sensitive to the inclusion or exclusion of transfers. Some Committee members also requested further explanation of the exclusion of outlier episodes from the MSPB measure. Not excluding outlier episodes results in MSPB measure scores which have a very high, statistically significant correlation of 0.95 with the MSPB measure scores that do exclude outlier episodes. This finding indicates that the MSPB measure is not very sensitive in aggregate to the outlier exclusion.

Acute-to-acute transfer cases are excluded from starting an MSPB episode, based on public comment through notice and comment rulemaking. Stakeholders expressed concern with attributing the episode to a hospital that did not treat the patient for the whole index hospitalization. They specifically expressed concern that attribution of an episode to a receiving hospital would disadvantage hospitals often called upon to receive transfers, because follow-up care may be received in a region outside the influence of the hospital receiving the transfer (42 CFR 51621).

Acumen's analysis shows that transfer episodes' observed spending is almost twice as expensive as non-transfer episodes' observed spending ($34,801 vs. $18,381). This is largely an artifact of the inpatient payment system, which pays more in total for a transfer than for one hospital stay. Transfer episodes account for 2 percent of total episodes. If transfer episodes were to be included in the measure, the attribution method would be especially important due to their high cost.

In the NQF Measure Submission Form, Acumen evaluated assigning transfers to either the transferring hospital or to the receiving hospital. This analysis found that small rural facilities are more likely to transfer their patients than are large urban facilities (3.7 percent vs. 1.5 percent). Thus, attributing a transfer episode only to the transferring hospital would disadvantage small rural facilities, while attributing the episode only to the receiving hospital would disadvantage large urban facilities (although the effect was of lower magnitude, due to a larger patient population). When transfer episodes are assigned to the receiving hospital, 90 percent of hospitals experience a change in their MSPB measure values of less than 3 percentage points, and 80 percent of hospitals experience a change in their MSPB measure values of less than 3 percentage points when transfer episodes are assigned to the transferring hospital.

To supplement these analyses and address the Committee's concerns, Acumen evaluated the impact of attributing transfer episodes to *both* the receiving and transferring hospitals. To

gauge the impact of this attribution, Acumen utilized two different strategies for assigning the transfer episodes. The first method assigns the episode to both hospitals equally. The second method assigns the episode to both hospitals, but weights the episode according to the percentage of total length of stay that occurred at each hospital. There is a high and statistically significant rank correlation between the MSPB measure excluding transfers versus the MSPB measure including transfers under either approach to attributing to both hospitals. When weighting the transfer episode equally for both hospitals, the rank correlation is 0.97 while weighting the transfer episode by length of stay (LOS) for each hospital, the rank correlation increases to 0.99. This indicates that the MSPB measure is not very sensitive in aggregate to the inclusion of transfers. For additional information regarding this analysis, please refer to the workbook titled "NQF_Transfers_Analysis_18JUN2013" attached with this memorandum.

Outliers are excluded from the MSPB measure calculation to avoid cases where a small number of high-cost or low-cost outliers have a disproportionate effect on a hospital's MSPB measure. In the NQF Measure Submission Form, Acumen evaluated the impact of top-coding and bottom-coding outlier episodes instead of excluding them, and found that the results were highly correlated with the original methodology of excluding outliers. Committee members asked what the results would look like if outlier episodes were fully included. Acumen performed this analysis and found a very high, statistically significant Spearman rank correlation of 0.95 between hospitals' MSPB measures calculated excluding outliers and hospitals' MSPB measures calculated including outliers. This high positive correlation indicates that the exclusion of outliers has very little effect on the ranking of hospitals in the MSPB measure.

Due to the importance of end-of-life care, some members of the Committee were also concerned that the exclusion of episodes where a beneficiary dies may be removing important information from the MSPB measure.CMS finalized this feature of the MSPB measure through notice and comment rulemaking, because episodes during which a beneficiary dies can be problematic in comparing to other episodes. Episodes in which a beneficiary died in the hospital have no post-discharge window at all, and post-discharge costs are the main driver of MSPB episode cost variation. In this case, costs that might have occurred if the beneficiary had not died are not observable. On the other hand, episodes in which the beneficiary dies towards the end of the 30 day post-discharge period often have very high expenses due to intensive end of life care. To avoid including episodes of care with incomplete costs, episodes during which a beneficiary dies are currently excluded from the MSPB measure calculation.

Based on analysis that demonstrates that episodes during which the beneficiary dies have higher observed spending that episodes during which the beneficiary lives ($22,364 vs. $18,966,

respectively), CMS will consider including episodes where a beneficiary dies in the MSPB measure calculation for future measure refinement.[12]

4. Risk Adjustment Methodologies

Summary: For the MSPB measure, the look-back period is the timeframe during which hierarchical condition categories (HCCs) are gathered from claims data and used for risk adjustment. In the NQF Measure Submission Form, Acumen showed that switching from a 90-day look-back period to a 365 day look-back did not materially change the R^2 of the regression (0.4621 to 0.4601), indicating that the 90 day look-back is performing just as well as a 365 day look-back at predicting MSPB spending.

Committee members suggested also including Present on Admission diagnoses from the index admission in the risk adjustment model. Doing so results in an MSPB measure that has a very high correlation of 0.99 with the original MSPB measure, indicating no practical impact. The R^2 increases slightly from 0.45 to 0.46, although this is not likely to be meaningful. Our conclusion is that including Present on Admission diagnoses has very little impact on the risk adjustment model performance or final MSPB measure scores.

Additionally, some Committee members expressed concern with the R^2 results presented in the NQF Measure Submission Form, which prompted Acumen to look more closely at them and realize that it is not possible to calculate within-decile R^2. Acumen examined average predicted and observed spending in each decile and found that they are similar within each decile; observed spending also increases monotonically from lower deciles to higher deciles. The difference in cost between lower deciles and higher deciles is substantial. Together, these facts show that the model is discriminating well between high cost and low cost episodes, and that it is predicting cost well throughout the distribution.

One Committee member also asked if Acumen could test using the natural log of spending, instead of the level. We find that using the natural log worsens the fit of the model. The R^2 is 0.41 when observed episode cost is the dependent variable in the regression. On the other hand, the R^2 is 0.39 when the natural log of observed episode cost is the dependent variable in the regression, indicating a worse fit. Together, these finding supports the validity of the risk adjustment methodology.

Some members of the Committee were concerned that the 90-day "look-back period" in the MSPB measure risk adjustment methodology is too short to sufficiently capture beneficiaries' comorbidities. Acumen previously presented a comparison with using a one year

[12] Note, however, that the same analysis demonstrates that MSPB episodes during which a beneficiary dies have lower risk-adjusted spending than episodes during which the beneficiary lives ($16, 411 vs. $18,817, respectively).

look-back, and found that the risk adjustment model had a slightly lower R^2 with a one year look-back than with a 90-day look-back (0.4621 to 0.4601).

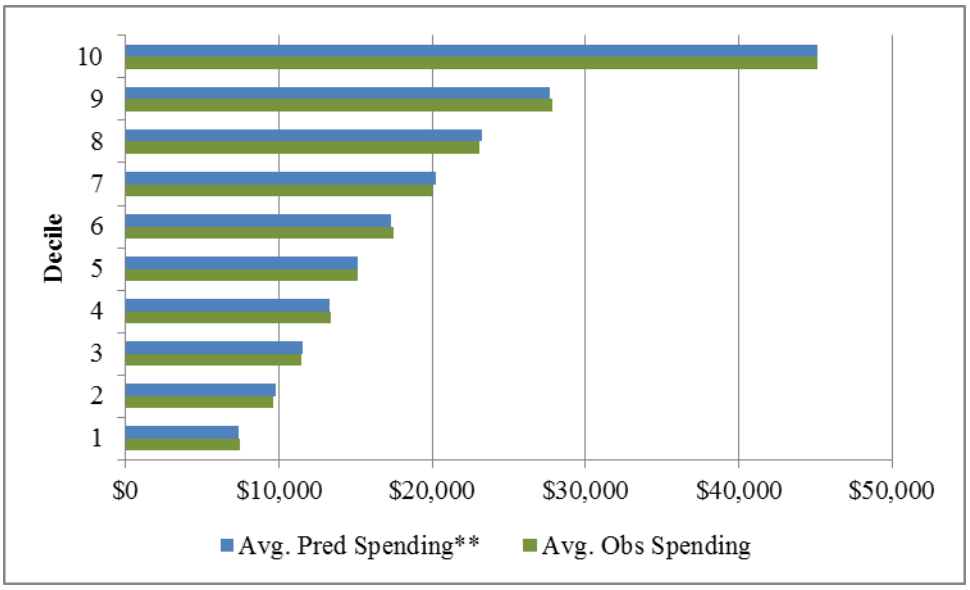
In addition, Committee members suggested including all diagnoses present on admission in the risk adjustment methodology. To address this concern, Acumen compared the correlation of hospitals' MSPB measures calculated under the current risk adjustment methodology against hospitals' MSPB measures calculated when the risk adjustment methodology includes all diagnoses present on admission, as indicated by the Present on Admission (POA) indicators on the index admission claim. Using January 2012 – December 2012 Medicare Parts A and B claims data, Acumen found a very high correlation of 0.99 between hospitals' MSPB measures calculated using the current risk adjustment methodology and hospitals' MSPB measures calculated including all diagnoses present on admission in the risk adjustment methodology. This very high positive correlation indicates that the current risk adjustment methodology is very similar to one which includes diagnoses present on admission, indicating that the exclusion of these POA diagnoses does not adversely affect the measure's validity. Including all diagnoses present on admission in the risk adjustment methodology slightly increases the R^2 of the model from 0.45 to 0.46, although this is likely not statistically meaningful. Our conclusion is that including Present on Admission diagnoses has very little impact on the risk adjustment model performance or final MSPB scores and if anything, including them could potentially subject the measure to "gaming," as hospitals control the diagnoses on the claim. For additional information regarding this analysis, please refer to the workbook titled "NQF_Including_Present_On_Admission_Dgn_22MAY2013" attached with this memorandum.

In the NQF Measure Submission Form, Acumen also calculated the distribution of episode spending and R^2 by decile (where deciles are defined by the predicted cost) to examine the model's ability to predict costs throughout the distribution.[13] Some Committee members expressed concern with the R^2 results presented in the NQF Measure Submission Form, which prompted Acumen to look more closely at them and realize that it is not possible to calculate R^2 within deciles. After further research and consultation, Acumen has conducted a more meaningful decile analysis that focuses on whether the average predicted spending in each decile closely fits the average observed spending in the decile and on whether observed spending increases monotonically with each decile (since the deciles are defined based on predicted cost). As can be seen in Table 8 and Figure 1 below, both of these criteria hold, indicating that the MSPB risk adjustment methodology is discriminating well and is predicting episode cost well throughout the distribution. For additional information regarding this analysis, please refer to the workbook titled "NQF_Model_Calibration_13JUN2013" attached with this memorandum.

[13] Please refer to Table A: Distribution of Spending and R-Squared by Decile (Includes Outlier Episodes) in the NQF Measure Submission Form.

Additionally, Acumen has examined the effect of risk adjustment by calculating the 90/10 ratio of MSPB episode cost both before and after risk adjustment. Risk-adjusting episode costs decreases the 90/10 ratio by almost 50 percent from 6.6 to 3.4. Table 9 presents these results as well as episode-level cost percentiles. This analysis shows that the risk adjustment is performing well in reducing the variation in observed spending. Both the decile analysis and 90/10 ratio analysis support the validity of the risk adjustment methodology.

Figure 1: Distribution of Average Observed and Average Predicted Spending by Decile



****Predicted Spending is the predicted value from the regression**

Table 8: Distribution of Average Observed and Average Predicted Spending by Decile

Decile	Episode Count	Avg. Obs Spending	Avg. Pred Spending**
1	446,268	$7,442	$7,365
2	446,234	$9,607	$9,763
3	446,197	$11,472	$11,506
4	446,234	$13,379	$13,276
5	446,260	$15,164	$15,114
6	446,205	$17,452	$17,350
7	446,512	$20,047	$20,226
8	445,951	$23,108	$23,237
9	446,130	$27,830	$27,631
10	446,339	$45,115	$45,148
TOTAL	**4,462,330**	**$19,062**	**$19,062**

****Predicted Spending is the predicted value from the regression**

Table 9: Episode-Level Observed and Risk-Adjusted Costs

Cost	90/10 Ratio	Standard Deviation	Percentile of Cost						
			10	25	50	75	90	95	99
Observed	6.6	$14,543	$5,632	$7,787	$13,773	$24,866	$37,225	$45,742	$65,746
Risk-Adjusted by Ratio	3.4	$10,775	$9,241	$11,410	$15,066	$21,617	$31,822	$39,760	$58,767
Risk-Adjusted by Residual	3.3	$9,495	$9,469	$12,686	$15,837	$21,595	$31,608	$38,139	$51,541

One Committee member also asked if Acumen could test using the natural log of spending, instead of the level. We find that using the natural log worsens the fit of the model. The R^2 is 0.41 when observed episode cost is the dependent variable in the regression. On the other hand, the R^2 is 0.39 when the natural log of observed episode cost is the dependent variable in the regression, indicating a worse fit. This indicates that the relationship of the independent variables, most of which are binary, with observed cost is not well described as logarithmic.

5. Cost variation by Setting of Care

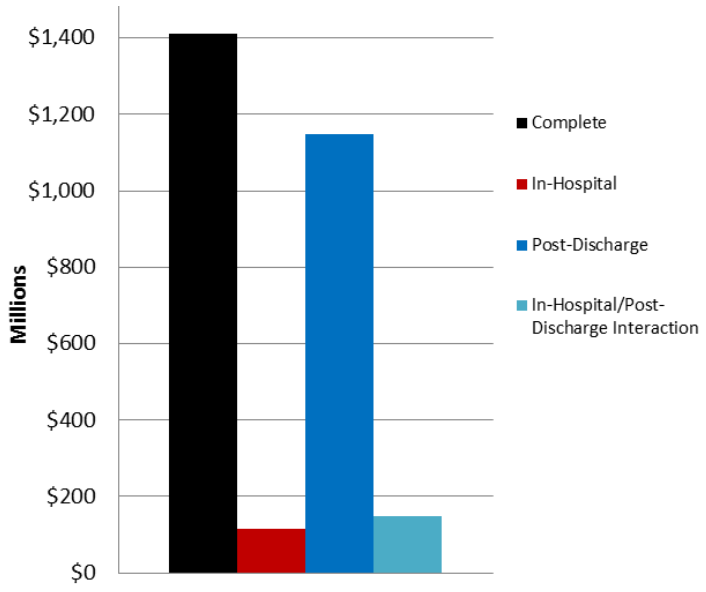
Summary: The Committee was interested in whether variation in the MSPB measure is largely driven by post-discharge spending. Acumen has divided the total variance in MSPB risk-adjusted spending into index admission costs, and post-discharge costs. As expected, variance in risk-adjusted post-discharge cost accounts for the large majority of the total risk-adjusted cost variance.

Several Committee members were interested in how much of the variation in the MSPB measure is driven by variation in post-discharge spending versus index admission spending. In the NQF Measure Submission Form and supplementary materials, Acumen only addressed this question for non-risk-adjusted cost. However, the risk adjustment controls for MS-DRG, which substantially changes the relative variation between index admission spending and post-discharge spending. To address this, Acumen has broken down the total variance in risk-adjusted cost by time period (index admission vs. post-discharge). One would expect risk-adjusted episode cost to be strongly driven by post-discharge cost, since the MS-DRG of the index admission almost completely determines the inpatient payment, leaving only variation in index admission professional fees.

Acumen found (as expected) that the variance in post-discharge costs makes up a larger portion of total variance than index admission costs do. Figure 2 shows that post-discharge costs account for approximately 80 percent of total episode cost variance, while index admission ("in-hospital") costs account for approximately 9 percent of total episode cost variance. Decomposing post-discharge variance by setting also reveals that IP Hospital and SNF costs are the main drivers of post-discharge variance.

Acumen would like to emphasize that this finding is for risk-adjusted costs both during the hospitalization and post hospital discharge. The risk adjustment model predicts a certain level of post-discharge spending based upon the beneficiary's prior health history and MS-DRG. This analysis shows that of the cost variance left over after this risk adjustment, most of it is driven by post-discharge spending. For additional information regarding this analysis, please refer to the workbook titled "NQF_Variance_Analysis_10JUN2013" attached with this memorandum.

Figure 2: Variance Decomposition of Risk-Adjusted Episode Cost by In-Hospital vs. Post-Discharge



Measure Reliability

1. Test/Retest Analysis

Summary: Acumen split all beneficiaries into two random, non-overlapping samples. A quintile stability analysis shows a highly stable relationship between the samples. Seventy (70) percent of the top quintile in one sample remains in the top quintile in the other, while 90 percent of the top quintile in one sample remains in the top two quintiles. For a completely random measure (i.e., a measure that is unreliable from one sample to another), these figures would be 20 percent and 40 percent, respectively. In addition, the Spearman rank correlation between the two samples is 0.84 and statistically significant. Both of these analyses indicate that the MSPB measure is highly reliable.

Some members of the Committee expressed concern with the results from Acumen's "test-retest" analysis in which Acumen examined the correlation and quintile rank stability between a hospital's MSPB score measured using two non-overlapping random samples. Specifically, some members of the Committee were concerned with the result that approximately 30 percent of hospitals in the lowest-spending quintile in one sample move to a different quintile in the next sample and that approximately 30 percent of hospitals in the highest-spending quintile in one sample move to a different quintile in the next sample.

The quintile stability analysis between the two random samples showed that over 70 percent of hospitals in the lowest-spending quintile in one sample are in the lowest-spending quintile in the other sample; similarly, over 70 percent of hospitals in the highest-spending quintile in one sample are in the highest-spending quintile in the other sample. If the MSPB measure were completely random (i.e., unreliable from one sample to another), this number would be expected to be only 20 percent. In addition, over 90 percent of hospitals in the highest-spending quintile in one sample are in the top two quintiles in the other sample. This is a highly stable result for quintile stability analyses. In addition, Acumen found that the Spearman rank correlation across samples is 0.84 and statistically significant. This large correlation coefficient indicates a highly stable measure. This conclusion is supported by Carlos Alzola, the NQF's statistical consultant, who said during the NQF Cost and Resource Use Steering Committee Meeting that the Spearman rank correlation was more than sufficient and satisfied him with respect to reliability.

2. Period of Performance Analysis

Summary: Acumen tested using both an 8 month period of performance and a 12 month period of performance to calculate the MSPB measure and found that the resulting sets of scores are highly correlated (0.97). This shows that the MSPB measure is reliable and robust to specification changes.

Acumen examined the correlation of hospitals' MSPB measures calculated using different length periods of performance. This analysis tests whether the measure is reliable by testing its sensitivity to the period of performance length. Using January 2012 – December 2012 Medicare Parts A and B claims data, Acumen compared hospitals' MSPB measures using a period of performance from January 1, 2012 to December 31, 2012 against hospitals' MSPB measures using a period of performance from May 1, 2012 to December 31, 2012. Acumen found that hospitals' MSPB measures with a 12 month period of performance exhibit a very strong, positive Spearman rank correlation of 0.97 with hospitals' MSPB measures with an 8 month period of performance, indicating that both periods of performance give hospitals similar MSPB measures. This reinforces Acumen's previous finding that an 8-month period of

performance is comparable to a full year of performance on the MSPB measure and supports the reliability of the measure with respect to a minimum period of performance length of 8 months.

APPENDIX

Appendix Tables A and B provide additional information on analyses discussed in the memorandum. Appendix Table A presents the correlation of the MSPB measure and several utilization measures constructed for various categories of medical services as well as a combined utilization of services category (see Measure Validity: 1. Correlation with Other Cost Measures). Appendix Table B, on the other hand, presents various stratifications of the MSPB measure by geographic location region, and teaching status (see Measure Validity: 2. Stratification by Characteristics). Appendix Table C supplements Appendix Table B by providing states located within each region breakdown in Appendix Table B.

Appendix Table A: Correlation Between MSPB Measures (May 2011- Dec 2011) and Utilization Measures

	Professional E&M Services	Procedures Services	Hospital Facilities Services		Emergency Services	Ancillary Services	Post-Acute		Other		Total
			Inpatient Setting	Outpatient Setting			IP & SN	HH	IP	Non IP	
	(units)	(units)	(util days)	(units)	(units)	(units)	(util days)	(count of claims)	(util days)	(units)	
Correlation Value	0.585	0.130	0.221	0.013	0.213	0.073	0.595	0.265	-0.012	0.146	0.224
P value	0.00	0.00	0.00	0.4582	0.00	0.00	0.00	0.00	0.4873	0.00	0.00

IP: Inpatient

SN: Skilled Nursing

HH: Home Health

E&M: Evaluation and Management

Appendix Table B: Impact Analysis, MSPB Breakdowns by Geographic Location, Region, and Teaching Status

	MSPB Measure			Average MSPB Amount	Average Spending Per Episode	Average Expected Spending Per Episode	# of Hospitals	%of Hospitals	
	Average	Minimum	Maximum					MSPB Measure ≥1	MSPB Measure <1
BY GEOGRAPHIC LOCATION:									
All Hospitals	**0.98**	**0.44**	**1.86**	**$17,998**	**$18,358**	**$18,358**	**3,369**	**42.4%**	**57.6%**
Large Urban	1.01	0.48	1.82	$18,526	$19,092	$18,773	1,326	57.4%	42.6%
Other Urban	0.98	0.44	1.86	$17,927	$18,342	$18,516	1,101	39.7%	60.3%
Rural Area	0.95	0.44	1.45	$17,336	$15,835	$16,450	942	24.3%	75.7%
BY REGION:									
New England	1.01	0.44	1.19	$18,568	$18,232	$17,836	143	71.3%	28.7%
Middle Atlantic	0.99	0.51	1.28	$18,143	$18,484	$18,270	384	51.0%	49.0%
South Atlantic	0.99	0.57	1.82	$18,085	$18,304	$18,285	540	41.3%	58.7%
East North Central	0.99	0.58	1.62	$18,138	$18,269	$18,258	514	47.9%	52.1%
East South Central	0.99	0.56	1.86	$18,128	$17,593	$17,737	321	39.3%	60.7%
West North Central	0.93	0.47	1.15	$17,050	$17,807	$18,604	268	16.8%	83.2%
West South Central	1.01	0.63	1.71	$18,520	$19,046	$18,505	551	56.6%	43.4%
Mountain	0.95	0.44	1.37	$17,357	$18,541	$18,843	237	25.7%	74.3%
Pacific	0.96	0.48	1.53	$17,561	$18,630	$19,040	411	28.2%	71.8%
Puerto Rico									
BY TEACHING STATUS:									
Non-Teaching	0.98	0.44	1.86	$17,920	$17,696	$17,715	2,351	39.9%	60.1%
Teaching	0.99	0.58	1.33	$18,177	$19,006	$18,988	1,018	48.1%	51.9%

Appendix Table C: States by Region

New England	Middle Atlantic	South Atlantic	East North Central	East South Central	West North Central	West South Central	Mountain	Pacific
Connecticut	Pennsylvania	Delaware	Illinois	Alabama	Iowa	Arkansas	Arizona	Alaska
Massachusetts	New Jersey	D.C.	Indiana	Kentucky	Kansas	Louisiana	Colorado	California
Maine	New York	Florida	Michigan	Mississippi	Minnesota	Oklahoma	Idaho	Hawaii
New Hampshire		Georgia	Ohio	Tennessee	Missouri	Texas	Montana	Oregon
Rhode Island		North Carolina	Wisconsin		Nebraska		Nevada	Washington
Vermont		South Carolina			North Dakota		New Mexico	
		Virginia			South Dakota		Utah	
		West Virginia					Wyoming	

MEMORANDUM

TO: Cost and Resource Use Steering Committee

National Quality Forum

FROM: Centers for Medicare & Medicaid Services and Mathematica Policy Research

DATE: June 27, 2013

REFERENCE: Responses to Cost and Resource Use Steering Committee Concerns regarding the Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service Beneficiaries (#2165)

We appreciate the opportunity to respond to the NQF Cost and Resource Use Steering Committee's comments regarding the Payment-Standardized Total Per Capita Cost Measure for Medicare Fee-for-Service (FFS) Beneficiaries (#2165). This whole-person cost measure assesses annual Medicare payments (payment-standardized and risk-adjusted) to encourage greater efficiency in the care furnished to Medicare FFS beneficiaries by physician groups. Beneficiaries are attributed to physician group practices, not individual physicians, that provide the plurality of services commonly provided by primary care physicians, because these physician groups are often in a good position to oversee the cumulative cost of their own services and the services that their beneficiaries receive from other providers.

CMS developed this measure to comply with Congress's directive for CMS to (a) develop a per capita measure of physician resource use (Section 1848(n)(1)(B)(ii) of the Social Security Act) and (b) use the measure (among others) to determine a value-based payment modifier (VBM) to be applied to physician group payment in such a way to encourage greater efficiency in the furnishing of services to Medicare FFS beneficiaries under the Medicare Physician Fee Schedule (Section 1848(p) of the Social Security Act). Congress also directed us to adjust this measure to account for variations in health status and other patient characteristics. In addition, CMS is required to provide this measure, along with other quality information, in confidential feedback reports (the Quality and Resource Use Reports [QRURs]) to physician group practices to assist them in better managing the resources involved with furnishing care to Medicare FFS beneficiaries.

In this memorandum, we describe the additional analyses we have performed to address the Steering Committee's concerns regarding measure construction issues related to attribution, risk adjustment, and exclusions, as well as the actionability of the measure. The additional data analyses described herein demonstrate that the measure is soundly constructed and a valid and reliable measure of per capita spending on Part A and Part B services for Medicare beneficiaries, thus not necessitating changes to the measure specifications at this time. Nonetheless, as CMS engages in future notice and comment rulemaking to apply the VBM to more groups of physicians (including groups of non-physicians such as nurse practitioners) and to individual physicians in upcoming years, it is committed to further examining these issues to better hone the measure to address the Steering Committee's concerns while complying with its statutory mandates. CMS also plans to continue to refine the QRURs that it shares with groups of physicians to better provide actionable information regarding costs to enable them to deliver high quality care at lower cost.

The remainder of this memorandum summarizes the major concerns raised by the Committee, results from additional testing we conducted to address these concerns, and our responses to Committee

concerns. The appendix presents the data source and results of measure testing on attribution, risk adjustment, and part-year beneficiary exclusions. In addition, we describe the further analysis we plan to perform later this summer to address the Committee's concerns.

Committee's Concerns and Responses

I. Attribution

Committee's First Concern: Lack of inclusion of nurse practitioners (NPs) and physician assistants (PAs), many of whom provide primary care, in the first step of the attribution methodology may attribute beneficiaries to groups that should not be responsible for the total cost of the beneficiary's care.

Response: Although adding NPs and PAs to the first step of attribution increases the number of beneficiaries attributed to group practices, more than 97 percent of attributed beneficiaries under the current attribution method are attributed to the same group when NPs and PAs are added in the first step of attribution.

The current attribution method for physician groups (not individual physicians) is based on the attribution rule used in the Medicare Shared Savings Program (MSSP), which is an Agency-wide two-step approach to attribution. Under Step 1, a Medicare beneficiary is attributed to a group if he/she received a plurality of services commonly provided by primary care physicians—hereafter referred to as primary care services PCS)—from a primary care physician (PCP) that is affiliated with the group. A PCP is defined as a physician practicing internal medicine, family practice, general practice, or geriatric medicine. Step 2 applies only if a beneficiary does not meet the criterion under Step 1. Under Step 2, a beneficiary is attributed to the group from whom he/she received at least the plurality of his/her PCS from non-PCP physicians, NPs, PAs, and clinical nurse specialists, including at least one PCS from a physician.

To explore the impact of excluding NPs and PAs from Step 1 of the attribution rule, we tested a modification of the MSSP attribution rule based on two key changes. First, we included NPs and PAs in the first step of the attribution rule, along with (as previously) family practitioners, general practitioners, geriatric medicine specialists, and internists. Second, we eliminated the requirement that a beneficiary must receive at least one PCS from a physician in the practice to which the beneficiary is attributed. In Exhibit I-1 in the Appendix, we show the number of beneficiaries attributed, the number of groups with attributed beneficiaries, the number of beneficiaries attributed via the first and second steps, and the mean percentage of PCS provided by the group who was attributed the beneficiary under both the MSSP and the revised attribution rules. As a result of this change, the number of beneficiaries and the number of groups to which beneficiaries were attributed increased slightly:

- The number of beneficiaries attributed to groups increased by 2.6 percent; the number of groups to which at least 20 beneficiaries were attributed increased by 3.4 percent (Exhibit I-1).

- Overall, over 97 percent of beneficiaries were attributed to the same group under both attribution rules. On average, about 2.3 percent of beneficiaries per group were attributed to different groups when NPs and PAs were added to the first step, meaning these beneficiaries received most of their PCS from these professionals rather than from PCPs. Thus, only a small proportion of beneficiaries and groups is affected by this change.

Consistent with statute, the Physician Feedback and Value-Based Payment Modifier Program measures and rewards the performance of *physicians*. As required by Medicare Improvement for Patients

and Providers Act of 2008, Public Law 110-275, and subsequently by the Affordable Care Act of 2010, Public Law 111–148, the Total Per Capita Cost Measure has focused on attributing beneficiaries to groups whose physicians—as opposed to NPs and PAs—provided PCS to its patients. However, we recognize the increasingly important contribution of NPs and PAs to primary care. Section 1848(p) of the Social Security Act provides that starting in 2017, CMS may extend the VBM to cover non-physician clinicians. In future rulemaking, CMS plans to consider including NPs and PAs—just as PCPs currently are included—in the first step of the attribution rule. In doing so, we will also examine ways to correctly identify, via claims, those NPs and PAs that are providing primary care from those practicing in subspecialty care settings.[1]

Committee's Second Concern: The attribution rule holds primary care providers accountable for all patient costs, including specialist costs.

Response: The Total Per Capita Cost Measure is a whole-person measure that captures total annual spending. As such, the MSSP attribution approach is oriented toward those groups of physicians who have the most influence over the patient's entire healthcare experience, and therefore, resource use.

In our 2011 sample, the average group accounted for a significant majority (68 percent, or approximately five out of seven visits for the typical beneficiary) of their attributed beneficiaries' evaluation and management visits in office, other outpatient, skilled nursing facility, and home settings over the course of the year. Given this high percentage, we believe that such a group will generally be in a good position to be responsible for overseeing or managing total annual costs per year. Nonetheless, we acknowledge the concern and will continue to monitor the proportion of services provided by the groups who are attributed beneficiaries versus other medical group practices going forward.

II. Risk Adjustment

Committee's Concern: The Total Per Capita Cost Measure includes dual eligibility status,[2] a socioeconomic indicator, in the risk adjustment model, whereas NQF recommends against adjusting for factors associated with socioeconomic status unless there is a clear clinical rationale for doing so.

Response: Dual eligibility status as a factor in the risk adjustment model conveys beneficiary clinical information not captured otherwise, in addition to socioeconomic status. There is clear evidence indicating that diagnostic information (the basis for CMS' Hierarchical Condition Category (CMS-HCC) model) is not sufficiently capturing all of the meaningful clinical differences between dual eligible and non-dual eligible beneficiaries. For example, Kautter et al. (2008) found statistically significant differences in the frailty of dual eligible versus non-dual eligible beneficiaries that predict differences in Medicare spending even after accounting for differences in diagnoses.[3] As frailty is not accounted for within the CMS-HCC model, the addition of dual eligibility status in the model can more accurately predict expenditures.

---

[1] The Number of Nurse Practitioners and Physician Assistants Practicing Primary Care in the United States: Primary Care Workforce Facts and Stats No. 2. October 2011. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/research/findings/factsheets/primary/pcwork2/index.html.

[2] Dual eligible beneficiaries are Medicare beneficiaries who are also eligible for Medicaid.

[3] Kautter, J., Ingber, M., & Pope, G. C. (2008). Medicare risk adjustment for the frail elderly. *Health Care Financing Review, 30*(2), pp. 83-93.

This observation reflects a more general finding that we confirmed with our 2011 sample, namely, that dual eligible beneficiaries are sicker than non-dual eligible beneficiaries along a large number of dimensions. Using the 70 HCC indicators used in our risk adjustment model, we summed the total number of HCC indicators for dual eligible and non-dual eligible beneficiaries so as to use the total number of HCC indicators as a proxy of health status. As shown in Exhibit II-1, dual eligible beneficiaries had twice the number of HCCs than their non-dual eligible counterparts (at the median), when comparing the groups based on the total number of HCCs. In addition, we examined the percentage of beneficiaries with each condition for dual versus non-dual beneficiaries. Exhibit II-2 shows that dual eligible beneficiaries were more likely to have 64 out of the 70 HCCs than non-dual eligible beneficiaries (p-value<0.001). The fact that dual eligible beneficiaries have a demonstrably different comorbidity profile along both measured and unmeasured dimensions of the CMS-HCC model is strong evidence in our view of the importance of explicitly accounting for them within the model, especially in the absence of credible direct controls for frailty and functional limitations.

Finally, we note that the inclusion of dual eligibility status satisfies Section 3003 of the Affordable Care Act, which explicitly calls for risk adjustment for the QRURs based on socioeconomic, demographic, and health status information.[4]

III. Exclusions Analysis

Committee's Concern: By excluding beneficiaries without a full year of Part A and Part B costs, costs that may have a significant effect on total per capita costs, such as those associated with end-of-life care, are excluded from the Total Per Capita Cost Measure.

Response: Including part-year beneficiaries who died during the year, were enrolled in Medicare Advantage for only part of the year, or were newly enrolled during the measurement year does not substantially change groups' per capita cost performance.

To examine the effects of the part-year exclusion criteria on the Total Per Capita Cost Measure, we included beneficiaries who died between January 1 and December 31 in the measurement year, were enrolled for part of the year in Medicare Advantage and part year in FFS, or were newly enrolled in Parts A and B during any time in the measurement year. For part-year beneficiaries, costs were annualized and weighted by the portion of 2011 that each beneficiary was enrolled in both Medicare Parts A and B.[5] We first examined the impact of such a change on the distribution of costs (Exhibit III-1). Next, we examined whether group rankings would change based on the inclusion of part-year beneficiaries in two ways: first, we looked at Spearman rank correlations (Exhibit III-2), and second, we compared quintile rankings (Exhibit III-3), between the full year model and the model with part–year beneficiaries included.

---

[4] Section 3003 (QRURs): ''(D) DATA ADJUSTMENT.—In preparing reports under this paragraph, the Secretary shall make appropriate adjustments, including adjustments—''(i) to account for differences in socioeconomic and demographic characteristics, ethnicity, and health status of individuals (such as to recognize that less healthy individuals may require more intensive interventions); and (ii) to eliminate the effect of geographic adjustments in payment rates (as described in subsection (e)).

[5] Costs are annualized by first weighting part-year beneficiaries' total costs by the fraction of months in the year they were enrolled (e.g., if a beneficiary is enrolled for six months, her costs are annualized by dividing her six months' total costs by 0.5). Annualized costs are then weighted by the same fraction for the calculation of per capita costs for groups. Annualizing costs is in effect the same as imputing costs for the months during which part-year beneficiaries were not enrolled and for which we do not have cost data.

Including part-year beneficiaries did little to affect the mean per capita costs, where the groups were located in the distribution of per capita costs, and their classification as a high or low performer. More specifically, Exhibit III-1 shows that while adding in all part-year beneficiaries increased mean per capita costs by 9.6 percent, it did not have a significant impact on the groups' comparative performance. Furthermore, as shown in Exhibit III-2, there was very little change in the ordering of group costs between the full-year only model and the model with part-year beneficiaries. Correlations were above 0.98 (p-value<0.0001). Lastly, Exhibit III-3 shows that upwards of 9 out of 10 groups representing the highest and lowest performers on this measure would remain as such with their attributed part-year beneficiaries included.[6]

The annualization of costs requires imputation of costs for some medical group practices, which may not be advisable. For example, if death is occurring more systematically among some group practices compared to others, such as at medical group practices composed of a majority of geriatric medicine specialists, imputation of costs can exacerbate the effect of end-of-life costs. Thus, by avoiding imputation, we believe that the current measure—with the part-year exclusion criteria—does not systematically bias some groups over others and allows for fairer comparisons among groups.

IV. Actionablity of Group-Level Information

Committee's Concern: The group-level information included in the Quality and Resource Use Reports (QRURs) to date is not sufficiently actionable for individual providers, as there can be wide variation in performance among individual providers within groups.

Response: In the fall of 2013, based on performance year 2012, CMS will provide QRURs to groups of physicians with 25 or more eligible professionals. The QRURs contain the group practice's cost measure, as well as, benchmark information to help them identify how they compare to their peers. The QRURs also breakdown the measure into specific service costs (e.g., Evaluation and Management Services, Procedures, Hospitalizations, Emergency Services, Ancillary Services, Post-Acute Services), along with the percentage of patients receiving these services.

Beginning with the reports to be released in fall 2013, CMS will provide each group with information on beneficiaries attributed to the group. This drill-down detail will improve the actionability of the QRURs by enabling medical group practices to identify beneficiaries associated with higher resource use. For each beneficiary attributed to the group, the drill-down tables include the following information: (1) beneficiary demographic data (gender, date of birth, HCC risk score profile, whether the beneficiary died during measurement year); (2) date of last professional service claim filed by the group; (3) number of primary care services provided by the group; (4) percentage of primary care services billed by the group; (5) percentage contribution of specific services to total costs that match the summary information provided to the group; (6) list of all hospital admissions and whether the admission was a readmission, was associated with an ambulatory-care sensitive condition, and other diagnostic and discharge information; and (7) presence of specific chronic conditions among the group's patient population: diabetes, chronic obstructive pulmonary disease, heart failure, and coronary artery disease. We developed this list of information based on feedback we received from recipients of previous QRURs

[6] Cohen's Kappa Statistic, which evaluates the agreement across categorical variables (i.e., inter-rater agreement), was also computed. The Kappa was equal to 0.79, which indicates substantial agreement across performance categories.

and believe that this additional drill-down capacity will increase the actionability of these reports for individual providers within groups. We are committed to further refining the information in the future reports based on the feedback we receive.

Conclusions and Future Analyses

After investigation of the key concerns of the Steering Committee, we are maintaining the Total Per Capita Cost Measure as currently specified but will consider, through future notice and comment rulemaking, whether to include NPs and PAs in the first step of attribution when the VBM is extended to other non-physician clinicians. Absent a model to capture frailty and other more subtle aspects of health status, dual eligibility status is included in the risk adjustment model as a proxy for important clinically relevant information that is not currently represented in the model. Furthermore, while adding in part-year beneficiaries makes a modest impact on the distribution of per capita costs, it has an inconsistent effect on groups who would be attributed these beneficiaries. Thus, to ensure a stable cohort of patients and fairness for comparison, we will maintain our requirement of 12-months of Parts A and B enrollment. Because we want to provide actionable information to the medical group practices receiving QRURs and affected by the VBM, we intend to include service- and beneficiary- level detail for PY2012 to these groups.

The Steering Committee also expressed concern regarding the lack of Part D pharmacy data in the Total Per Capita Cost measure. Including outpatient prescription data in a total per capita resource use measure is an important component. We are currently exploring ways to report Part D pharmacy resource use separately in future QRURs and Agency-wide approaches to address this broader issue.

Finally, we are also conducting further analysis that will be ready in mid-August 2013 to address the Steering Committee's concerns regarding the lack of inclusion of Federally Qualified Health Centers and Rural Health Clinics in the attribution approach by examining whether beneficiaries who are attributed to medical group practices should actually be attributed to these providers and to further examine the relationship between dual eligibility status and resource use. We believe these analyses will be instructive as we move forward with this measure.

APPENDIX

ADDITIONAL INFORMATION FROM ANALYSIS

A. Data Source

All testing was conducted on the sample used in the original analysis as part of our measure submission package: medical group practices, identified by Taxpayer Identification Number (TIN), that satisfied the following criteria in 2011: (1) at least 25 eligible professionals (EPs) billed Medicare under the group's TIN; (2) at least 20 beneficiaries were attributed to the medical group practice; and (3) the medical group practice was located in California, Illinois, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, or Wisconsin.[7] Testing of the measure is based on Medicare Parts A and B claims and enrollment data for 2011, and CMS' Hierarchical Condition Category (HCC) risk scores that are used in risk adjustment.

B. Additional Measure Testing Results

I. Attribution Testing Results

In our original submission to the NQF, the Medicare Shared Saving Program (MSSP) attribution rule was applied. Under this rule, beneficiaries are attributed to groups in the first step if they received a plurality of primary care services (PCS)[8] from primary care physicians (PCP)—defined as physicians practicing internal medicine, family practice, general practice, or geriatric medicine— affiliated with these groups; if they do not meet the criteria for the first step, they may be attributed via the second step to the group with a non-PCP physician from whom they received at least one PCS and with affiliated NPs, PAs, and Clinical Nurse Specialists who provided the plurality of their PCS. We modified the rule (referred below as "Revised Attribution Rule") by allowing beneficiaries to be attributed to groups with Nurse Practitioners (NPs) and Physician Assistants (PAs) who provided the beneficiaries' plurality of primary care services and by removing the requirement that a beneficiary see a primary care physician to be attributed via the first step.

Looking at both attribution rules, we examined the number of beneficiaries attributed, the number of groups with attributed beneficiaries, the number of beneficiaries attributed via the first and second steps, and the mean percentage of PCS provided by the group who was attributed the beneficiary (Exhibit I-1).

---

[7] We determined which state a medical professional practiced in based on a plurality of carrier claims, using the state indicator field (e.g., a professional with 50 carrier claims in California, 40 in Kansas, and 20 in Michigan would be assigned to California). At the group level, the group was assigned to a state based on the most common state among affiliated medical professionals. Note that there could be medical group practices in the nine states that have 25 or more eligible professionals (EPs) nationally but fewer than 25 EPs in one of the nine states; these groups were excluded from the final sample for this analysis.

[8] Primary care services in the MSSP attribution rule are defined as: CPT codes: 99201–99205 (Office or other outpatient visits for the evaluation and management of a new patient); 99211–99215 (Office or other outpatient visit for the evaluation and management of an established patient); 99304–99306 (Initial nursing facility care, per day, for the evaluation and management of a patient); 99307–99310 (Subsequent nursing facility care, per day, for the evaluation and management of a patient); 99315–99316, 99318 (Nursing facility discharge day management); 99318 (Evaluation and management of a patient involving an annual nursing facility assessment); 99324–99328 (Domiciliary or rest home visit for the evaluation and management of a new patient); 99334–99337 (Domiciliary or rest home visit for the evaluation and management of an established patient); 99339–99340 (Individual physician supervision of a patient (patient not present) in home, domiciliary, or rest home); 99341–99345 (Home visit for the evaluation and management of a new patient).

As presented in Exhibit I-1, the revised attribution rule results in a modest increase of 2.55 percent in the number of beneficiaries who are attributed to groups and in a 3.36 percent increase in the number of groups with attributed beneficiaries.

Exhibit I-1. Comparison of Medicare Shared Savings Program Attribution Rule and Revised Attribution Rule for Groups with At Least 25 Eligible Professionals and At Least 20 Attributed Beneficiaries, 2011

	MSSP[a] Attribution rule	Revised Attribution Rule[b]	Percent Change
Total number of Beneficiaries	2,648,490	2,716,061	+2.55%
Total Number of Groups[c]	804	831	+3.36%
Mean number of attributed beneficiaries per group	3,294	3,268	-0.78%[e]
Number of beneficiaries attributed via the first step	2,259,785	2,409,665	+6.63%
Number of beneficiaries attributed via the second step	388,705	306,396	-21.18%
Mean percentage of PCS[d] provided by the group that is attributed the beneficiary per group	68.34	67.86	-0.71%[e]

Source: Medicare fee-for-service (FFS) claims data, January to December 2011.
Note: Medical groups only include those with at least 25 EPs and at least 20 attributed beneficiaries practicing in California, Illinois, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, or Wisconsin in 2011. Groups are identified by their taxpayer identification numbers (TINs).

[a] MSSP: Medicare Shared Savings Program.

[b] We revised the MSSP attribution rule by allowing beneficiaries to be attributed to groups with Nurse Practitioners and Physician Assistants who provided the beneficiaries' plurality of primary care services and by removing the requirement that a beneficiary see a primary care physician to be attributed via the first step.

[c] The total number of groups under the revised attribution rule may not include all groups that were included under the MSSP attribution rule.

[d] PCS: Primary Care Services

[e] The mean percentages decrease under the revised rule because the revised attribution rule results in a higher percentage increase in the number of groups with attributed beneficiaries (3.36 percent) than the number of beneficiaries who are attributed (2.55 percent); thus, there are more beneficiaries attributed over a larger number of groups. With a higher denominator under the revised rule, mean percentages are lower under the revised rule compared to the MSSP rule.

II. Risk Adjustment: Dual Eligibility Status Testing

Using the HCC indicator file from CMS based on performance year 2010 (as the risk adjustment model is a prospective model that is based on prior year risk scores), we matched the beneficiaries in our 2011 sample to the beneficiaries in the HCC indicator file. As the HCC indicators are 0-1 indicators indicating the presence of the diagnoses represented by each HCC, we summed the HCC indicators for dual eligible and non-dual eligible beneficiaries. The total number of HCC indicators, therefore, serves as a proxy of a beneficiary's health status. Exhibit II-1 shows the distribution of the number of HCC indicators, by dual eligibility status. In addition in Exhibit II-2, we examined the percentage of dual eligible and non-dual eligible beneficiaries who had each HCC indicator. The table also presents that ratio of the percentage of dual eligible beneficiaries to the percentage of non-dual eligible beneficiaries with a given condition. A ratio greater than one indicates that dual eligible beneficiaries are more likely than non-dual eligible beneficiaries to have a specific condition.

Dual eligible beneficiaries have twice the number of HCCs as non-dual eligible beneficiaries (at the median), which means that dual eligible beneficiaries are more likely to have greater comorbidities than non-dual eligible beneficiaries (Exhibit II-1). Similarly, when looking at each HCC indicator, dual eligible beneficiaries are significantly more like to have the vast majority of the HCC indicators (64 out of 70). In some cases, dual eligible beneficiaries are upwards of 10 percent more likely to have certain HCCs, such as HIV/AIDS (11 percent more likely) and Schizophrenia (18 percent more likely, Exhibit II-2).

Exhibit II-1. Distribution of Beneficiaries' Total Number of HCC Indicators, by Dual Eligibility Status

Dual Eligibility Status		Distribution of the Total Number of HCC Indicators[a]						
	Mean	Min	5%	25%	50%	75%	95%	Max
Dual Eligible Beneficiaries	2.11	0	0	1	2	3	7	27
Non-Dual Eligible Beneficiaries	1.46	0	0	0	1	2	5	22

Source: Medicare FFS claims data, January to December 2011.

Note: The total number of beneficiaries that are dual eligible is equal to 565,383 and non-dual eligible equal to 2,087,259 (pre risk adjustment). These are beneficiaries attributed to medical group practice (N=802) with at least 25EPs and at least 20 attributed beneficiaries practicing in California, Illinois, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, or Wisconsin in 2011. Groups are identified by their taxpayer identification numbers (TINs).

[a] The HCC Flag Indicators are 0-1 indicators, with 0 indicating that the condition is not present and 1 indicating the condition is present. For each beneficiary, the total number of HCC indicators was computed for this analysis. The maximum number of HCC indicators that a beneficiary can have is 70. The beneficiary sample is that used in the original NQF submission.

Exhibit II-2. Comparison of the Percentage of Dual Eligible and Non-Dual Eligible Beneficiaries Who Are Indicated As Having the HCC by HCC, 2011

HCC Indicator[a]	Percent of Beneficiaries with HCC Condition: Dual Eligibles (N= 565,383)	Percent of Beneficiaries with HCC Condition: Non-Dual Eligibles (N= 2,087,259)	Ratio of Dual Eligible Percentage to Non-Dual Eligible Percentage [b, c]
HCC1: HIV/AIDS	1.15	0.10	11.3
HCC2: Septicemia/Shock	2.27	1.00	2.3
HCC5: Opportunistic Infections	0.38	0.23	1.6
HCC7: Metastatic Cancer and Acute Leukemia	0.94	1.09	0.9
HCC8: Lung, Upper Digestive Tract, and Other Severe Cancers	0.77	0.87	0.9
HCC9: Lymphatic, Head and Neck, Brain, and Other Major Cancers	1.44	1.93	0.7
HCC10: Breast, Prostate, Colorectal and Other Cancers and Tumors	4.90	9.18	0.5
HCC15: Diabetes with Renal or Peripheral Circulatory Manifestation	5.11	2.83	1.8
HCC16: Diabetes with Neurologic or Other Specified Manifestation	4.73	2.76	1.7
HCC17: Diabetes with Acute Complications	0.16	0.08	1.9
HCC18: Diabetes with Ophthalmologic or Unspecified Manifestation	2.01	1.67	1.2
HCC19: Diabetes without Complication	17.67	15.63	1.1
HCC21: Protein-Calorie Malnutrition	1.89	0.81	2.3
HCC25: End-Stage Liver Disease	0.57	0.21	2.7
HCC26: Cirrhosis of Liver	0.68	0.30	2.3
HCC27: Chronic Hepatitis	1.17	0.23	5.2
HCC31: Intestinal Obstruction/Perforation	1.96	1.35	1.5
HCC32: Pancreatic Disease	1.70	1.09	1.6
HCC33: Inflammatory Bowel Disease	0.90	0.86	1.0
HCC37: Bone/Joint/Muscle Infections/Necrosis	1.34	0.69	1.9
HCC38: Rheumatoid Arthritis and Inflammatory Connective Tissue Disease	5.34	5.00	1.1
HCC44: Severe Hematological Disorders	1.07	0.77	1.4
HCC45: Disorders of Immunity	1.03	0.82	1.3
HCC51: Drug/Alcohol Psychosis	1.20	0.36	3.3
HCC52: Drug/Alcohol Dependence	2.73	0.49	5.5
HCC54: Schizophrenia	6.45	0.36	17.8
HCC55: Major Depressive, Bipolar, and Paranoid Disorders	12.77	3.67	3.5
HCC67: Quadriplegia, Other Extensive Paralysis	0.95	0.14	6.8
HCC68: Paraplegia	0.63	0.12	5.1
HCC69: Spinal Cord Disorders/ Injuries	0.88	0.45	2.0
HCC70: Muscular Dystrophy	0.18	0.04	4.6
HCC71: Polyneuropathy	7.20	5.28	1.4
HCC72: Multiple Sclerosis	1.17	0.47	2.5
HCC73: Parkinson's and Huntington's Diseases	1.62	1.42	1.1
HCC74: Seizure Disorders and Convulsions	7.67	1.72	4.5
HCC75: Coma, Brain Compression/Anoxic Damage	0.33	0.11	3.0
HCC77: Respirator Dependence/Tracheostomy Status	0.38	0.11	3.5
HCC78: Respiratory Arrest	0.05	0.02	2.4
HCC79: Cardio-Respiratory Failure and Shock	4.85	3.02	1.6
HCC80: Congestive Heart Failure	13.62	10.41	1.3
HCC81: Acute Myocardial Infarction	1.08	0.91	1.2
HCC82: Unstable Angina and Other Acute Ischemic Heart Disease	1.92	1.62	1.2
HCC83: Angina Pectoris/Old Myocardial Infarction	4.08	3.95	1.0
HCC92: Specified Heart Arrhythmias	10.17	13.77	0.7
HCC95: Cerebral Hemorrhage	0.47	0.36	1.3
HCC96: Ischemic or Unspecified Stroke1,498	3.79	2.66	1.4
HCC100: Hemiplegia/Hemiparesis	2.02	0.78	2.6
HCC101: Cerebral Palsy and Other Paralytic Syndromes	1.09	0.13	8.3
HCC104: Vascular Disease with Complications	2.47	1.87	1.3
HCC105: Vascular Disease	14.32	11.64	1.2
HCC107: Cystic Fibrosis	0.06	0.02	3.7
HCC108: Chronic Obstructive Pulmonary Disease	15.86	10.22	1.6
HCC111: Aspiration and Specified Bacterial Pneumonias	1.23	0.49	2.5
HCC112: Pneumococcal Pneumonia, Empyema, Lung Abscess	0.32	0.21	1.5
HCC119: Proliferative Diabetic Retinopathy and Vitreous Hemorrhage	1.25	0.69	1.8
HCC130: Dialysis Status	1.68	0.42	4.0
HCC131: Renal Failure	11.32	9.34	1.2
HCC132: Nephritis	0.29	0.19	1.5
HCC148: Decubitus Ulcer of Skin	2.07	0.74	2.8

HCC Indicator[a]	Percent of Beneficiaries with HCC Condition		Ratio of Dual Eligible Percentage to Non-Dual Eligible Percentage [b, c]
	Dual Eligibles (N= 565,383)	Non-Dual Eligibles (N= 2,087,259)	
HCC149: Chronic Ulcer of Skin, Except Decubitus	2.78	1.98	1.4
HCC150: Extensive Third-Degree Burns	0.01	0.00	2.1
HCC154: Severe Head Injury	0.02	0.01	2.7
HCC155: Major Head Injury	1.06	0.46	2.3
HCC157: Vertebral Fractures w/o Spinal Cord Injury	1.05	1.08	1.0
HCC158: Hip Fracture/Dislocation	1.28	1.05	1.2
HCC161: Traumatic Amputation	0.19	0.07	2.9
HCC164: Major Complications of Medical Care and Trauma	4.28	3.00	1.4
HCC174: Major Organ Transplant Status	0.56	0.29	1.9
HCC176: Artificial Openings for Feeding or Elimination	1.38	0.55	2.5
HCC177:Amputation Status, Lower Limb/Amputation Complications	0.60	0.20	3.1

Source: Medicare FFS claims data, January to December 2011.

Note: The total number of beneficiaries that are dual eligible is equal to 565,383 and non-dual eligible equal to 2,087,259 (pre risk adjustment). These are beneficiaries attributed to medical group practice (N=802) with at least 25EPs and at least 20 attributed beneficiaries practicing in California, Illinois, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, or Wisconsin in 2011. Groups are identified by their taxpayer identification numbers (TINs).

[a] The HCC Flag Indicators are 0-1 indicators, with 0 indicating that the condition is not present and 1 indicating the condition is present. For each beneficiary, the total number of HCC indicators was computed for this analysis. The maximum number of HCC indicators that a beneficiary can have is 70. The beneficiary sample is that used in the original NQF submission.

[b] For example, dual eligible beneficiaries were 11.3 times more likely to have HCC1 than non-dual eligible beneficiaries in 2011.

[c] All differences were statistically significant at the 1% significance level with the exception of Vertebral Fractures w/o Spinal Cord Injury.

III. Exclusions Analysis

The Total Per Capita Cost Measure as currently specified excludes beneficiaries who are not fully and continuously enrolled in Medicare FFS Parts A and B during the measurement year. Specifically, a beneficiary is excluded from the sample of beneficiaries if between January and December of the measurement year, they were indicated as having died, had partial year enrollment in Medicare Advantage, or were newly enrolled during performance year 2011. We built on the analysis of 802 groups with at least 25 eligible professionals and at least 20 attributed beneficiaries who were the basis for our original NQF analysis. Thus, all changes presented in per capita costs and the number of attributed beneficiaries are based on the same sample of 802 groups.

We first examined the impact of such a change on the distribution of costs (Exhibit III-1). Then, we examined whether groups rankings would change based on the inclusion of part year beneficiaries in two ways: first, we looked at Spearman rank correlations (Exhibit III-2) and second, we compared quintile rankings (Exhibit III-3), between the full year model and the model with part years included.

Exhibit III-1 shows that the full-year sample plus all part-year beneficiaries have mean per capita costs that are about 9.6 percent higher than mean per capita costs for the full-year sample only. However, the rank ordering of the groups' per capita costs between the full-year only model and the model with part-year beneficiaries were highly correlated, above 0.98 (p-value<0.0001; Exhibit III-2). Additionally, upwards of 9 out of 10 groups representing the highest and lowest performers on this measure would remain as such with their attributed part-year beneficiaries included (Appendix Exhibit III-3). For example, 93.7 percent of groups in the lowest quintile of per capita costs in the full-year only sample remain there when part-year beneficiaries' costs are added. Similarly, 90.6 percent of groups in the highest quintile remain there when part-year beneficiaries are included. Thus, the vast majority of high and low performers stay in the same position relative to their peer medical group practices when part-year beneficiaries are included.

Exhibit III-1. Number of Attributed Beneficiaries when Part-Year Beneficiaries Are Included for Groups with At Least 25 Eligible Professionals and At Least 20 Attributed Beneficiaries, 2011

Sample	Total Number of Attributed Beneficiaries	Mean Number of Attributed Beneficiaries	Percent Change in Attributed Beneficiaries (Comparison to Full-Year Only Sample)	Mean Per Capita Costs	Percent Change (Comparison to Full-Year Only Sample)
Full-Year Only	2,619,719	3,266	-	$10,602	-
Full-Year + Those Who Died[a]	2,725,880	3,399	4.05%	$11,738	10.71%
Full-Year + Those in Medicare Advantage[b]	2,685,787	3,349	2.52%	$10,553	-0.47%
Full-Year + New Enrollees[c]	2,772,363	3,457	5.83%	$10,656	0.51%
Full-Year + All Part-Year Enrollees[d]	2,926,092	3,648	11.69%	$11,622	9.62%

Source: Medicare FFS claims data, January to December 2011.

Note: The total number of beneficiaries that are dual eligible is equal to 565,383 and non-dual eligible equal to 2,087,259 (pre risk adjustment). These are beneficiaries attributed to medical group practice (N=802) with at least 25EPs and at least 20 attributed beneficiaries practicing in California, Illinois, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, or Wisconsin in 2011. Groups are identified by their taxpayer identification numbers (TINs).

[a] Beneficiaries who died between January 1 and December 31 during 2011.

[b] Beneficiaries with partial year enrollment in Medicare Parts A and B as well as partial year enrollment in Medicare Advantage during 2011.

[c] Beneficiaries who were new enrollees in Medicare Parts A and B during 2011.

[d] All part-year beneficiaries include those who died, were enrolled in Medicare Advantage for part of the year, and who were new enrollees in 2011.

Exhibit III-2. Spearman Rank Correlations between Full-Year Only Sample and Full-Year Sample with Part-Year Beneficiaries[a], By Exclusion Criteria, 2011

Full-Year Only	Full-Year Only	Full-Year + Those Who Died	Full-Year + Those in Medicare Advantage	Full-Year + New Enrollees	Full-Year + All Part-Year Enrollees
Correlation	1.000	0.981 (<.0001)	0.994 (<.0001)	0.994 (<.0001)	0.977 (<.0001)

Source: Medicare FFS claims data, January to December 2011.

Note: The exclusion criteria are not mutually exclusive. These are beneficiaries attributed to medical group practice (N=802) with at least 25 EPs and at least 20 attributed beneficiaries practicing in California, Illinois, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, or Wisconsin in 2011. Groups are identified by their taxpayer identification numbers (TINs). All part-year beneficiaries here include those who died, were enrolled in Medicare Advantage, and those who were newly enrolled. P-values shown in parentheses.

[a] Part-year beneficiaries included beneficiaries who (1) were new enrollees during any point in 2011, (2) were enrolled in Medicare Advantage during any point in 2011 but also had partial year enrollment in Parts A and B; and/or (3) died between January 1 and December 31, 2011.

Exhibit III-3. Comparison of Mean Per Capita Cost Quintiles When Part-Year [a] Beneficiaries are Included

Rankings when Part-Year Beneficiaries are **Excluded**	Rankings when Part-Year Beneficiaries are **Included** (Total Number and Percent of Groups in Each Quintile)					Total
	Lowest Cost Quintile	Second Quintile	Third Quintile	Fourth Quintile	Highest Cost Quintile	
Lowest Cost Quintile	151 (93.7%)	9 (5.6%)	0 (0.0%)	1 (0.6%)	0 (0.0%)	161 (20.0%)
Second Quintile	10 (6.3%)	126 (78.8%)	21 (13.1%)	3 (1.9%)	0 (0.0%)	160 (20.0%)
Third Quintile	0 (0.0%)	24 (14.9%)	120 (74.5%)	16 (9.9%)	1 (0.6%)	161 (20.0%)
Fourth Quintile	0 (0.0%)	1 (0.6%)	19 (11.9%)	126 (78.8%)	14 (8.8%)	160 (20.0%)
Highest Cost Quintile	0 (0.0%)	0 (0.0%)	1 (0.6%)	14 (8.8%)	145 (90.6%)	160 (20.0%)
Total	161 (20.0%)	160 (20.0%)	161 (20.0%)	160 (20.0%)	160 (20.0%)	802 (100%)
Kappa Statistic						0.79

Source: Medicare FFS claims data, January to December 2011.

Note: These are beneficiaries attributed to medical group practice (N=802) with at least 25 EPs and at least 20 attributed beneficiaries practicing in California, Illinois, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, or Wisconsin in 2011. Groups are identified by their taxpayer identification numbers (TINs).

[a] Part-year beneficiaries included beneficiaries who (1) were new enrollees during any point in 2011, (2) were enrolled in Medicare Advantage during any point in 2011 but also had partial year enrollment in Parts A and B; and/or (3) died between January 1 and December 31, 2011.