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.

<u>Measure Reliability</u>

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

14

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).
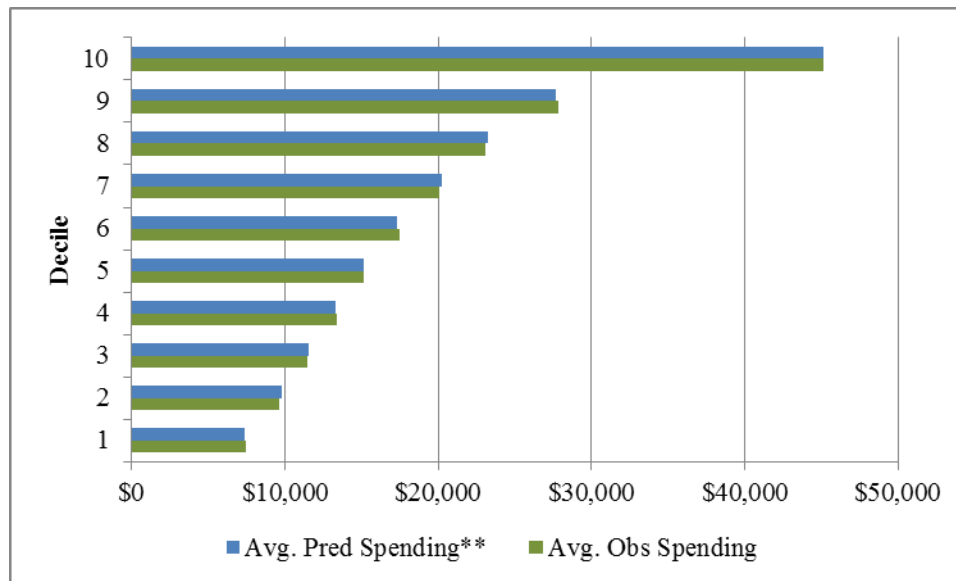
In addition, Committee members suggested including all diagnoses present on admission in the risk adjustment methodology. To address this concern, Acumen compared the correlation of hospitals' MSPB measures calculated under the current risk adjustment methodology against hospitals' MSPB measures calculated 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.
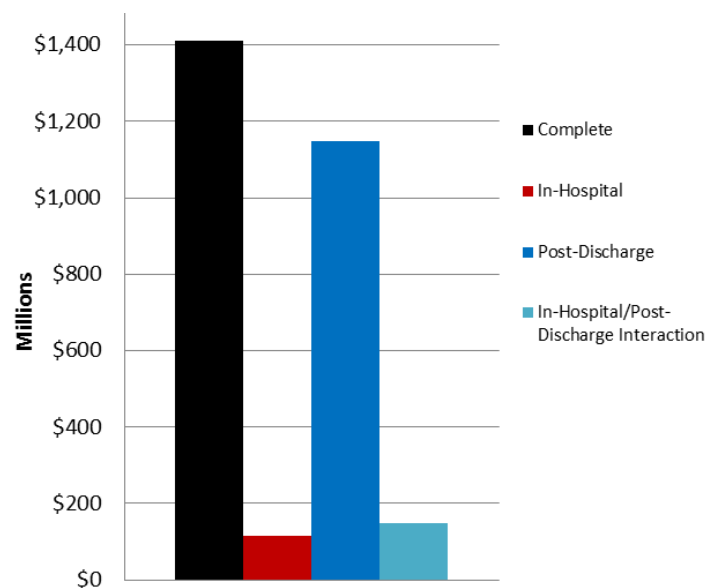
5. Cost variation by Setting of Care

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	