DEPARTMENT OF HEALTH & HUMAN SERVICES
Centers for Medicare & Medicaid Services
7500 Security Boulevard, Mail Stop S3-02-01
Baltimore, Maryland 21244-1850

MEMORANDUM

DATE: July 18, 2012

TO: Users of CMS' Hospital Value-Based Purchasing Program Website at http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing

FROM: Center for Medicare and
Center for Clinical Standards & Quality
Centers for Medicare & Medicaid Services

SUBJECT: Results of Reliability Analysis about Medicare Spending per Beneficiary Measure

The Centers for Medicare & Medicaid Services (CMS) is making public the analysis supporting the reliability of the Medicare Spending per Beneficiary measure, designed to evaluate hospitals' efficiency.

The analysis follows as a memorandum to CMS' policy analyst, Kimberly Spalding Bush, on the proceeding pages. Questions about this analysis may be directed to HospitalVBP@cms.hhs.gov.

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

MEMORANDUM

TO: Kimberly Spalding Bush
Centers for Medicare & Medicaid Services

FROM: Acumen, LLC

DATE: July 18, 2012

REFERENCE: Medicare Spending per Beneficiary (MSPB) Measure Reliability Analysis

---

The Medicare Spending per Beneficiary (MSPB) Measure evaluates hospitals' efficiency relative to the efficiency of the median hospital. Specifically, the MSPB Measure assesses the cost to Medicare of services performed by hospitals and other healthcare providers during an MSPB episode, which includes all Part A and Part B claims whose discharge date falls between 3 days prior to an inpatient prospective payment system (IPPS) hospital admission (index admission) through 30 days post-hospital discharge.[1]

This memorandum assesses the reliability of the MSPB Measure in two ways:

1. Reliability Statistic: the extent to which variation in the MSPB Measure is due to variation in hospital episode spending, rather than random variation due to the sample of cases observed. Using this methodology, the overall reliability of the MSPB Measure at a minimum of 10 episodes is 0.951.
2. Quintile Rank Stability: the extent to which repeated measurements of a hospital's MSPB Measure value agree with each other. Using this approach, the Pearson correlation is 0.829.

The remainder of this memorandum describes the methods used in the MSPB Measure reliability analyses listed above and presents a more detailed look at the findings, respectively.

METHODS

The content below is divided into two parts. The first section describes how the reliability statistic is calculated, and the second explains how the hospital's quintile rank stability measure is calculated. Results were calculated using Medicare claims data during an MSPB performance period that spans from May 15, 2010 through February 14, 2011.

[1] For detailed MSPB Measure specifications, please visit: http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1228772057350.

Reliability Statistic

Under the first reliability assessment, we calculate the reliability statistic using the amount of within-hospital variability in spending per beneficiary, the across hospital variability in spending per beneficiary, and the number of discharges assigned to each hospital.[2] Specifically, the reliability for a specific hospital (R_j) is calculated as:

$$R_j = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_{wj}^2 / n_j} \tag{1}$$

where σ_b^2 is the variance between all hospitals, σ_{wj}^2 is the variance within a specific hospital *j*, and n_j is the number of episodes at hospital *j*.

The variance between all hospitals, σ_b^2, is, in turn, calculated as:

$$\sigma_b^2 = \frac{SS_b}{num\_ep} \tag{2}$$

where *num_ep* is the total number of episodes at all hospitals, and SS_b is the sum of squares between all hospitals. Specifically, SS_b can be expressed as:

$$SS_b = \sum_j (n_j \ (\text{mean}_j - \text{mean}_{oa})^2) \ . \tag{3}$$

In words, SS_b is the sum over all hospitals of the squared difference between the mean episode at a specific hospital *j* ($mean_j$) and the mean episode over all hospitals ($mean_{oa}$). SS_b provides a sense of how much hospitals differ from the average. This weighting better represents the true variation in spending per beneficiary episodes by counting smaller hospitals less, as hospitals with fewer episodes will have a smaller impact on the overall variance of the MSPB episode distribution.

In equation (3), $mean_j$ is calculated as:

$$\text{mean}_j = \sum_{ep \in \{j\}} \left(\frac{observed\_pmt_{ep}}{predicted\_pmt_{ep}} \times weight_{ep} \right) \tag{4}$$

where $observed\_pmt_{ep}$ is the observed episode payment, $predicted\_pmt_{ep}$ is the predicted episode payment, $ep \in \{j\}$ indicates all episodes in the set of episodes attributed to hospital *j*, and $weight_{ep}$ is defined as the predicted payment for an episode *ep* divided by the sum of all episode predicted payments for a specific hospital *j*:

$$weight_{ep} = \frac{predicted\_pmt_{ep}}{\sum_j predicted\_pmt_{ep}} \tag{5}$$

[2] The calculation of the reliability statistic follows the methodology outlined in: Mathematica, Inc. "Memorandum: Reporting Period and Reliability of AHRQ, CMS 30-Day and HAC Quality Measures – Revised." http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP_Measure_Reliability-.pdf.

On the other hand, $mean_{oa}$ is calculated as the sum over all hospitals and over all episodes within the hospitals of $observed\_pmt_{ep}$ divided by $predicted\_pmt_{ep}$ with this quotient weighted by both $weight_{ep}$ and n_j:

$$\text{mean}_{oa} = \frac{\sum_j \left(n_j \times \sum_{ep} \left(weight_{ep} \times \frac{observed\_pmt_{ep}}{predicted\_pmt_{ep}} \right) \right)}{num\_ep} \tag{6}$$

Returning to equation (1), the variance within a specific hospital j, σ_{wj}^2, is, in turn, calculated as:

$$\sigma_{wj}^2 = \sum_{ep \in \{j\}} \left(weight_{ep} \times \left(\frac{observed\_pmt_{ep}}{predicted\_pmt_{ep}} - \text{mean}_j \right)^2 \right). \tag{7}$$

The overall reliability ($R_{overall}$) is calculated as the variance between all hospitals divided by the variance between all hospitals plus the mean squared difference within all hospitals (MS_{within}):

$$R_{overall} = \frac{\sigma_b^2}{\sigma_b^2 + MS_{within}} \tag{8}$$

where MS_{within} is calculated as the sum of the variance within a specific hospital j over all hospitals divided by the total number of episodes:

$$MS_{within} = \frac{\sum_j \sigma_{wj}^2}{num\_ep}. \tag{9}$$

Quintile Rank Stability

Under the second reliability assessment, we calculate quintile rank stability in order to assess the extent to which assessments of a hospital using different sets of claims produces similar measures of hospital performance on the MSPB Measure. Specifically, our second approach to assessing reliability considers the extent to which assessments of a hospital using randomly selected subsets of patients produces similar measures of hospital performance. This methodology is analogous to a "test-retest" approach in which hospital performance is measured once using a random subset of patients and then measured again using a second subset that excludes the MSPB episodes chosen for the first sample. Specifically, we created subsets by randomly dividing hospitals with greater than or equal to 50 observations into equal halves. By comparing the correlation of a hospital's MSPB Measure value calculated using the two mutually exclusive samples, one can identify the relationship of a hospital's score across multiple samples. Using these samples, one can also calculate quintile rank stability across groups.

FINDINGS

Reliability Statistic

Overall reliability of the MSPB Measure for hospitals with a minimum of 10 episodes is 0.951.[3] This figure corresponds to the average reliability for a hospital with at least 10 episodes. Previous work by Yale University proposed that a reliability threshold of 0.4 is the lower limit of "moderate" reliability.[4] The overall reliability significantly exceeds this threshold.

Not only is the MSPB Measure's average reliability high, but most individual hospitals also have reliable MSPB scores. When examining all hospitals (including ones with less than 10 episodes), 96.9 percent of hospitals have a reliability score greater than or equal to 0.4; Table 1 below provides a breakdown of different reliability levels.

Table 1: Cumulative Distribution of Reliability Scores across Hospitals

Reliability	Episode Minimum	# Hospitals with Fewer Episodes	Percent of Hospitals
0.1	7.6	27	0.8%
0.2	17.0	45	1.3%
0.3	29.1	72	2.2%
0.4	45.3	102	3.1%
0.5	68.0	162	4.8%
0.6	102.0	216	6.5%
0.7	158.7	329	9.8%
0.8	272.0	602	18.0%
0.9	611.9	1268	37.9%

Quintile Rank Stability

Measuring reliability under the "test-retest" framework, the Pearson correlation for a hospital across samples is 0.829 and the Spearman rank correlation for a hospital across samples is 0.835; both these values suggest high reliability.[5] In addition, Table 2 shows that when randomly dividing hospitals with greater than or equal to 50 observations into equal halves to calculate quintile rank stability across groups, over 70 percent of hospitals in the bottom quintile in one sample are in the bottom quintile in the next; similarly, over 70 percent of hospitals in the top quintile in one sample are in the top quintile in the next, suggesting high reliability.

[3] MSPB overall reliability increases by 0.0002 when the minimum number of episodes increases from 10 to 25.
[4] See Mathematica, Inc. memorandum cited above: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP_Measure_Reliability-.pdf
[5] The Pearson correlations measure the linear dependence of a hospital's of MSPB scores in the Group A sample with its MSPB score in the Group B sample. The Spearman rank correlations, on the other hand, measure the statistical dependence of the hospital's MSPB *rank* in Group A relative to their *rank* in Group B.

Table 2: Quintile Rank Stability Across Groups*

		Quintile in Group B				
		1	**2**	**3**	**4**	**5**
Quintile in Group A	**1**	72%	19%	6%	2%	1%
	2	21%	51%	21%	6%	2%
	3	4%	22%	45%	23%	5%
	4	2%	6%	23%	50%	19%
	5	1%	2%	6%	19%	73%

* To calculate the quintile in each group, hospitals with ≥ 50 observations had their observations randomly divided into equal halves.