

Medicare Spending Per Beneficiary Clinician Measure

Measure Justification Form

June 2019



Contents

1.0	Introduction	4
1.1	Project Title and Overview	4
1.2	Measure Name	4
1.3	Type of Measure	4
2.0	Importance.....	5
2.1	Evidence to Support the Measure Focus	5
2.1.1	Measure Description	5
2.1.2	Evidence for Measure Focus	5
2.2	Performance Gap	6
2.2.1	Rationale	6
2.2.2	Performance Scores.....	6
3.0	Scientific Acceptability	7
3.1	Data Sample Description	7
3.1.1	Type of Data Used for Testing.....	7
3.1.2	Specific Dataset Used for Testing	7
3.1.3	Dates of the Data Used in Testing.....	7
3.1.4	Levels of Analysis Tested.....	7
3.1.5	Entities Included in the Testing and Analysis.....	7
3.1.6	Patient Cohort Included in the Testing and Analysis.....	7
3.1.7	Sample Differences	8
3.1.8	Social Risk Factors Included in Analysis	9
3.2	Reliability Testing.....	9
3.2.1	Level of Reliability Testing	9
3.2.2	Method of Reliability Testing	9
3.2.3	Statistical Results from Reliability Testing	10
3.2.4	Interpretation	11
3.3	Validity Testing	11
3.3.1	Level of Validity Testing.....	11
3.3.2	Method of Validity Testing	11
3.3.3	Statistical Results from Validity Testing	12
3.3.4	Interpretation	13
3.4	Exclusions Analysis	14
3.4.1	Method of Testing Exclusions.....	14
3.4.2	Statistical Results from Testing Exclusions.....	15
3.4.3	Interpretation	15
3.5	Risk Adjustment or Stratification	16
3.5.1	Method of Controlling for Differences	16
3.5.2	Conceptual, Clinical, and Statistical Methods	17
3.5.3	Conceptual Model of Impact of Social Risks.....	17
3.5.4	Statistical Results	17
3.5.5	Analyses and Interpretation in Selection of Social Risk Factors.....	18
3.5.6	Method for Statistical Model or Stratification Development.....	19
3.5.7	Statistical Risk Model Discrimination Statistics	19
3.5.8	Statistical Risk Model Calibration Statistics	20
3.5.9	Statistical Risk Model Calibration – Risk Decile.....	20
3.5.10	Interpretation.....	20
3.6	Identification of Meaningful Differences in Performance.....	20
3.6.1	Method	20

3.6.2 Statistical Results	21
3.6.3 Interpretation	21
3.7 Missing Data Analysis and Minimizing Bias.....	21
3.7.1 Method	21
3.7.2 Missing Data Analysis	22
3.7.3 Interpretation	22
4.0 Feasibility.....	23
4.1 Data Elements Generated as Byproduct of Care Processes	23
4.2 Electronic Sources	23
4.3 Data Collection Strategy	23
4.3.1 Data Collection Strategy Difficulties.....	23
5.0 Usability and Use	24
5.1 Use	24
5.1.1 Current and Planned Use	24
5.1.2 Feedback on the Measure and Development Process	24
5.2 Usability	27
5.2.1 Improvement	27
5.2.2 Unexpected Findings.....	27
5.2.3 Unexpected Benefits	27
6.0 Related and Competing Measures	28
6.1 Relation to Other Cost Measures	28
6.2 Harmonization.....	28
6.3 Competing Measures	28
Contact Information.....	29

1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the revised Medicare Spending Per Beneficiary (MSPB) clinician cost measure. The MJF is intended to provide detailed information about the testing conducted on this measure, and accompanies the Measure Methodology and Measure Codes List file, which together comprise the specifications for this cost measure.¹

1.1 Project Title and Overview

The Centers for Medicare & Medicaid Services (CMS) has contracted with Acumen, LLC to develop care episode and patient condition groups for use in cost measures to meet the requirements of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). The contract name is “MACRA Episode Groups and Cost Measures.” The contract number is HHSM-500-2013-13002I, Task Order HHSM-500-T0002.

1.2 Measure Name

Medicare Spending Per Beneficiary (MSPB) clinician

1.3 Type of Measure

Cost/Resource Use

¹ CMS, “Medicare Spending per Beneficiary clinician Measure Methodology,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-mspb-clinician-codes-list.zip>.
CMS, “Medicare Spending per Beneficiary clinician Measure Codes List,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-mspb-clinician-codes-list.zip>.

2.0 Importance

2.1 Evidence to Support the Measure Focus

2.1.1 Measure Description

The MSPB clinician measure assesses the cost to Medicare as a result of the services performed by an individual clinician during an MSPB clinician episode, which comprises the period immediately prior to, during, and following a patient's inpatient hospital stay.

The cost measure score is a clinician's average risk-adjusted cost across all episodes attributed to the clinician. This measure includes costs of certain services that are incurred during a beneficiary's care in the period 3 days prior to an initial hospitalization, known as the index admission, through 30 days after the discharge from this hospitalization. The costs of the services incurred during this period of time, also known as the episode window, are attributed to the clinician responsible for managing the beneficiary's care. A defined list of services that are unlikely to be influenced by the clinician's care decisions and that are considered clinically unrelated to the management of care is not included in the measure. The beneficiary populations eligible for the MSPB clinician measure include Medicare beneficiaries enrolled in Medicare Parts A and B during the performance period.

2.1.2 Evidence for Measure Focus

Policymakers contend that an estimated 80 percent of overall health care costs are attributable to decisions made by clinicians.² However, these same clinicians are often unaware of how their care decisions influence the overall costs of care. One of the goals for using cost measures is to help inform clinicians on the costs attributable to their decision-making, as well as the total cost of their patient's care. A cost measure offers opportunity for improvement if clinicians can exercise influence on a significant share of costs during the episode, or if lower spending and better care quality can be achieved through changes in clinical practice.

As health expenditures continue to increase in the United States, the MSPB clinician measure is an important means of measuring Medicare spending, which is the largest single purchaser of health care in the United States. According to the National Health Expenditure Accounts, total health care spending is estimated to have increased by 4.6 percent in 2017, reaching \$3.5 trillion (CMS, 2018), and spending for Medicare, which is still predominantly paid on a fee-for-service (FFS) basis, grew by 3.6 percent, reaching \$672.1 billion.³ In 2016, Medicare FFS paid \$183 billion for approximately 10 million Medicare inpatient admissions and 200 million outpatient services, which reflects a 2.3 percent increase in hospital spending per FFS beneficiary between 2015 and 2016.⁴

Successfully establishing payment models under MIPS can have significant impacts on reducing costs and making care more affordable⁵. Population-based measures serve an essential role in measuring the cost of care and can serve as a transparent tool to control and curb growing health care costs.

² Fred, Herbert L. "Cutting the Cost of Health Care: The Physician's Role." Texas Heart Institute Journal, vol. 43, no. 1, 2016, pp. 4-6.

³ "National Health Expenditure Projections, 2017-2026." US Centers for Medicare & Medicaid Services, 2018.

⁴ Report to the Congress: Medicare Payment Policy." MedPAC, 2018.

⁵ Data Book: Health Care Spending and the Medicare Program." MedPAC, 2017.

2.2 Performance Gap

2.2.1 Rationale

Given that the inpatient hospital setting is such an important contributor to overall Medicare spending, gauging the efficacy of this spending requires measuring the cost performance of clinicians providing care at hospitals. The MSPB clinician measure provides valuable context for such progress in efficiency by measuring costs of care from a holistic perspective at the beneficiary level.

As background to the MSPB clinician measure, a version of the measure (referred to as the “MSPB measure”) has been part of the Merit-based Incentive Payment System (MIPS) cost performance category since the 2017 MIPS performance period. Prior to this use in MIPS, CMS used the MSPB measure in the Value Modifier Program and reported it in annual Quality and Resource Use Reports (QRURs) until MACRA ended the Value Modifier Program.

The current MSPB measure went through re-evaluation in 2018 to address stakeholder feedback received from prior public comment periods. A summary of the differences between the revised and the current MSPB measure can be found in Appendix A of the Cost Measure Methodology for the revised MSPB clinician measure on the CMS MACRA Feedback webpage.⁶

2.2.2 Performance Scores

Performance scores are provided for 20,853 clinician group practices (identified by Tax Identification Number [TIN]) and 127,529 practitioners (identified by combination of TIN and National Provider Identifier [NPI]). These counts represent attributed clinicians and clinician groups billing Part B Physician/Supplier claims under a MIPS eligible clinician specialty, and do not reflect other MIPS eligibility criteria (e.g., Advanced APM participation). This table uses a testing volume threshold of 35 episodes.

Table 1: Distribution of Performance Scores

Metric	TIN	TIN-NPI
Mean score	\$18,838	\$19,366
Standard deviation	\$1,785	\$1,854
Score IQR	\$2,061	\$2,296
Score percentile		
10 th	\$16,798	\$17,149
20 th	\$17,485	\$17,875
30 th	\$17,962	\$18,395
40 th	\$18,345	\$18,837
50 th	\$18,701	\$19,262
60 th	\$19,093	\$19,696
70 th	\$19,536	\$20,176
80 th	\$20,098	\$20,757
90 th	\$20,982	\$21,659

⁶ CMS, “Medicare Spending per Beneficiary clinician Measure Codes List,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-mspb-clinician-codes-list.zip>.

3.0 Scientific Acceptability

3.1 Data Sample Description

3.1.1 Type of Data Used for Testing

Medicare administrative claims, Long-Term Minimum data set (MDS), enrollment database (EDB), and Common Medicare Environment (CME)

3.1.2 Specific Dataset Used for Testing

The MSPB clinician measure uses Medicare Part A and Part B claims data maintained by CMS. Part A and B claims data are used to build episodes of care, calculate episode costs, and construct risk adjusters. Data from the EDB is used to determine beneficiary-level exclusions and supplemental risk adjusters, specifically Medicare Parts A, B, and C enrollment, primary payer, disability status, end-stage renal disease (ESRD), beneficiary birth dates, and beneficiary death dates. The risk adjustment model also accounts for expected differences in payment for services provided to beneficiaries in long-term care based on the data from the MDS. Specifically, the MDS is used to create the long-term care indicator variable in risk adjustment.

For measure testing, data from the American Census, American Community Survey (ACS), and CME are used in analyses evaluating social risk factors in risk adjustment.

3.1.3 Dates of the Data Used in Testing

The measurement period includes MSPB clinician episodes ending from January 1, 2017 through December 31, 2017.

3.1.4 Levels of Analysis Tested

Individual clinician (identified by combination of TIN and NPI) and clinician group/practice (identified by TIN).

3.1.5 Entities Included in the Testing and Analysis

20,853 clinician group practices and 127,529 practitioners were included in the analyses. Clinicians and clinician groups were included in testing if they were attributed 35 or more MSPB clinician episodes during the measurement period. Episodes from all 50 States and D.C. with an index admission in the acute inpatient setting were included.

3.1.6 Patient Cohort Included in the Testing and Analysis

4,250,421 Medicare beneficiaries (from 6,217,677 episodes) were included in TIN level measure testing, and 3,904,021 beneficiaries (from 5,650,783 episodes) were included in TIN-NPI level measure testing.

The beneficiaries included in the MSPB clinician measure calculation are enrolled in Medicare Parts A and B (but not Part C) and have had an admission to an acute care hospital setting. Beneficiaries and their episodes were included in the sample if they met a set of inclusion criteria (listed below) meant to ensure completeness of data and to focus the measure on a clinically homogeneous cohort of patients

The inclusion criteria are:

- The beneficiary has Medicare as their primary payer for any time during the episode window and 90-day lookback period prior to the episode start day used for risk adjustment.

- The beneficiary was continuously enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window, and was not enrolled in Medicare Part C for any time during this duration.
- The index admission of the episode was in an acute IP facility located in the United States.
- The beneficiary date of birth is not missing.
- The beneficiary death date did not occur before the episode end date.
- The index admission for the episode occurred in either a subsection (d) hospital paid under the Inpatient Prospective Payment System (IPPS) or in an acute hospital in Maryland⁷
- The discharge of the episode index admission did not occur in the last 30 days of the measurement period⁸
- The index admission for the episode was not involved in an acute-to-acute hospital transfer (i.e. the admission does not end in a hospital transfer or does not begin because of a hospital transfer)
- The claim for the index admission indicated a positive actual or standardized payment.

To determine whether the MSPB clinician measure's inclusion criteria distort patient characteristics on episodes, we produced and analyzed distributions of patient characteristics (age, race, sex, dual eligibility status, income, unemployment, hierarchical condition categories [HCCs]) for (i) episodes with inclusion criteria, (ii) episodes without inclusion criteria, (iii) beneficiaries with inclusion criteria, and (iv) beneficiaries without inclusion criteria.

This analysis shows that the MSPB clinician measure's inclusion criteria have only a minimal effect on the percentage of beneficiaries of any particular demographic. The difference between beneficiaries being included or not included in the measure is between -1.4 and 1.4 percentage points across each of the characteristics in the analysis. To illustrate, the percentage of beneficiaries aged 65 to 69 without applying the inclusion criteria is 18.0 percent while the percentage of beneficiaries aged 65 to 69 with applying the inclusion criteria is 17.2 percent. The difference in the percentage of beneficiaries for race with and without the inclusion criteria is near 0 percentage point difference for the most categories, and is between a 0.3 and 0.4 percentage points for two categories. The breakdown of male and female beneficiaries is similar when comparing the use of inclusion criteria, with 54.4 percent without inclusion criteria and 55.8 percent with inclusion criteria for female. These results indicate that there is minimal shift in patient characteristics after application of the inclusion criteria listed above.

3.1.7 Sample Differences

n/a

⁷ Subsection (d), which covers hospitals in the 50 states and D.C., does not include psychiatric hospitals, rehabilitation hospitals, hospitals whose inpatients are predominantly under 18 years old, hospitals whose average inpatient length of stay exceeds 25 days, and hospitals involved extensively in treatment for or research on cancer.

⁸ The measurement period is a pre-defined and static calendar year performance period. Episodes ending during the measurement period are included in the calculation of the revised MSPB clinician measure.

3.1.8 Social Risk Factors Included in Analysis

The social risk factors analyzed were variables from the ACS, EDB, and CME. All ACS variables are at the Census Block Group level. Social risk variables analyzed include the following:

- Income (ACS)
 - Low Income: median income < 33rd percentile nationally
 - Medium Income: median income in the interval spanning the 33rd percentile to the 66th percentile nationally
 - High Income: median income > 66th percentile
- Education (ACS)
 - Education < High School: when % with < high school education is the highest for a given Census Block Group
 - Education = High School: when % with only high school is the highest
 - Education > High School: when % with > high school is the highest
- Employment (ACS)
 - Unemployment Rate > 10%
 - Unemployment Rate <= 10%
- Race (EDB)
 - Asian, Black, Hispanic, North American Native, White, and Other
- Sex (EDB)
 - Female, male
- Dual status (CME)
 - Full dual, partial dual, non-dual

3.2 Reliability Testing

3.2.1 Level of Reliability Testing

The following levels of reliability were tested: critical data elements used in the measure and performance measure score (e.g., signal-to-noise analysis).

3.2.2 Method of Reliability Testing

Data Element Reliability

The MSPB clinician measure is constructed using CMS claims data, as described in section 3.1.2. CMS has implemented several auditing programs to assess overall claims code accuracy, ensure appropriate billing, and recoup any overpayments. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in this measure, including diagnosis and procedure codes and other elements that are consequential to payment. Specifically, CMS works with Zone Program Integrity Contractors, and formerly Program Safeguard Contractors, to ensure program integrity; the agency also uses Recovery Audit Contractors to identify and correct for underpayments and overpayments.

CMS also uses the Comprehensive Error Rate Testing (CERT) Program to ensure that Medicare payments are correct in accordance with coverage, coding, and billing rules. Between 2005 and 2017, CERT estimates that proper payment, which includes payments that met Medicare coverage, coding, and billing rules, ranged from 87.3 to 96.4 percent of total payments

each year.⁹ The FY 2018 Medicare FFS program proper payment rate was 91.9 percent.¹⁰ CMS continues to perform successful corrective actions and give providers additional education to ensure accurate billing.

To ensure claims completeness and inclusion of any corrections, the measure was developed and tested using data with a three month claims run-out from the end of the measurement period.

Measure Reliability

Measure reliability is the degree to which repeated measurements of the same entity agree with each other. For measures of clinician performance, the measured entity is the TIN or TIN-NPI, and reliability is the extent to which repeated measurements of the TIN or TIN-NPI give similar results. To estimate measure reliability, we used a signal-to-noise analysis.

This approach seeks to determine the extent to which variation in the measure is due to true, underlying clinician performance, rather than random variation (i.e., statistical noise) within clinicians due to the sample of cases observed. To achieve this, we calculate reliability scores as:

$$R_j = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_{w_j}^2}$$

Where:

$\sigma_{w_j}^2$ is the within-group variance of the mean measure score of clinician j

σ_b^2 is the between-group variance of clinicians within the episode group

That is, reliability is calculated as the ratio of between-group variance of clinicians to the sum of between-group variance and within-group variance. Reliability closer to a value of one indicates that the between-group variance is relatively large compared to the within-group variance, which suggests that the measure is effectively capturing the systematic differences between the clinician and their peer cohort.

3.2.3 Statistical Results from Reliability Testing

Measure Reliability

As shown in the table below, 100 percent of TINs and TIN-NPIs have reliability greater than or equal to 0.4 at the 30, and 35-episode volume thresholds. Mean reliability increases with increasing volume thresholds. At a testing volume threshold of at least 35 episodes, the mean reliability for TINs is 0.77 and for TIN-NPIs is 0.69.

Table 2: Reliability Results at Various Volume Thresholds

Volume Threshold (# episodes)	TIN		TIN-NPI	
	Mean Reliability	% ≥ 0.4	Mean Reliability	% ≥ 0.4
20	0.70	100.0%	0.62	93.0%
30	0.75	100.0%	0.67	100.0%
35	0.77	100.0%	0.69	100.0%

⁹ Comprehensive Error Rate Testing (CERT) Program. “Appendices Medicare Fee-for-Service 2018 Improper Payments Report”. Table A6. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/CERT/Downloads/2018MedicareFFSSupplementalImproperPaymentData.pdf>.

¹⁰ Ibid.

3.2.4 Interpretation

Measure Reliability

Overall reliability of the MSPB clinician measure at a volume threshold of 35 episodes is moderately high for TINs and TIN-NPIs due to the large number of episodes attributed to clinicians. CMS generally considers 0.4 as the threshold for ‘moderate’ reliability and 0.7 as the threshold for ‘moderate to high’ reliability.¹¹

While higher volume thresholds yield even higher reliability results, it is at the cost of further reducing the number of clinicians and clinician groups able to receive a measure score.

3.3 Validity Testing

3.3.1 Level of Validity Testing

We conducted performance measure score validity testing, which included systematic assessment of face validity and empirical validity testing.

3.3.2 Method of Validity Testing

Face Validity

The MSPB clinician measure was developed through a structured process for gathering detailed input from recognized clinician experts on the measure. These expert panels were convened to methodically assess the extent to which the measure: (i) captured what it was intended to capture and (ii) differentiated between provider performance. Experts in this clinical area evaluated specifications in an iterative process to ensure that each aspect of the measure (e.g., attribution logic or service exclusion rules) was intentionally capturing only the costs of care within the reasonable influence of the attributed clinician (i.e., the ability of the measure score to differentiate good from poor performance).

In re-evaluating this measure, Acumen incorporated input from (i) three technical expert panel (TEP) meetings in 2017 and 2018, (ii) the MSPB Service Refinement workgroup, which convened in the summer of 2018, and (iii) stakeholder feedback from national field testing.

The TEP comprised 19 members from diverse backgrounds, including clinicians, healthcare providers, academia, and patient advocacy organizations, affiliated with 27 specialty societies. The MSPB Service Refinement workgroup was composed of 25 members, affiliated with 21 specialty societies, including the American Medical Association, Society of General Internal Medicine, Society of Thoracic Surgeons, and the American Academy of Family Physicians.

In the first meeting in August 2017, the TEP provided high-level guidance and initial input on direction of refinements, considering prior public comments and specifically focusing on attribution and service refinements. During the May 2018 meeting, the TEP provided input on specific approaches to refining attribution methodology and creating services exclusion logic. During the November 2018 meeting, the TEP reviewed feedback received on the measure from field testing and did not provide recommendations for further refinements.

The TEP recommended the creation of a targeted MSPB Service Refinement workgroup to provide detailed clinical input on service assignment rules. During two service refinement webinars in June and July 2018, the MSPB Service Refinement workgroup discussed empirical analyses and used their clinical expertise to assist with developing service assignment

¹¹ 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.

exclusions for the measure, with exclusions specific to Major Diagnostic Category (MDC) groupings. Input was gathered in a structured manner including the use of a polling process requiring greater than 60 percent consensus.

In addition, the national field testing feedback period in October and November 2018 offered all stakeholders an opportunity to review and provide input on draft measure specifications and measure feedback reports for attributed clinicians and clinician groups. During this period, 148,382 field test reports (20,852 for TINs and 127,530 for TIN-NPIs) were available for download and review for the MSPB clinician measure revised in 2018.

Empirical Validity Testing

We undertook two approaches to estimate the measure's validity. In the first approach, we evaluated the empirical validity of the MSPB clinician measure by examining differences in risk-adjusted cost for known indicators of resource or service utilization based on a literature review, specifically with downstream acute readmission and with post-acute care (PAC) service utilization. For this analysis, we compared the ratio of observed to expected cost (henceforth called the "O/E cost ratio") for MSPB clinician episodes with and without readmissions and with or without PAC services utilization. This analysis sought to confirm the expectation that the MSPB clinician measure captures variation in service utilization.

In the second approach, we evaluated how different types of cost impact risk-adjusted measure scores. Certain services or costs included in the MSPB clinician measure were classified into clinically coherent groups of services, called "clinical themes." The MSPB clinician measure categories of service are:

- **Acute Inpatient Service**, including acute inpatient hospital index admission, and services billed by the TIN/other TINs during index hospitalization
- **Inpatient Readmissions**, including acute inpatient hospitalization following the index admission and the related services billed by the TIN/other TINs
- **Post-Acute Care (PAC)**, including home health (HH), skilled nursing facility (SNF), and inpatient rehabilitation or long-term care facility (IRF/LTCH)
- **Emergency Services Not Included in a Hospital Admission**, including emergency E&M services; procedures; laboratory, pathology, and other tests; and imaging services.
- **Outpatient Evaluation and Management Services, Procedure, and Therapy (excluding emergency department)**, including physical, occupational, or speech and language pathology therapy; E&M services, major procedures; anesthesia, and ambulatory/minor procedures.

As with the first analysis for validity, the aim of this analysis was to determine whether the measure is capturing variation in provider cost in the manner intended and expected. To measure this, we calculated the Pearson correlation between the cost of each clinical theme and the overall risk-adjusted cost for an episode.

We expected that the readmission service category would have the highest correlation with risk-adjusted episode cost, as readmissions are likely associated with high cost even after accounting for beneficiary characteristics. Similarly we expected that the PAC: SNF service category would have a high correlation with risk-adjusted episode cost as well, since SNF services are frequently provided to beneficiaries after hospitalizations and tend to be less cost efficient than other PAC services (e.g., home health).

3.3.3 Statistical Results from Validity Testing

For the first analysis of validity, the mean O/E cost ratio for all episodes is 1.00. The mean O/E cost ratio for episodes with downstream acute readmission is 1.58, compared with 0.91 for episodes without downstream acute readmission. The mean O/E cost ratio for episodes with

PAC is 1.20, while for episodes without PAC is 0.80. Table 3 presents additional details on the O/E cost ratios for the various episode types.

Table 3: Distribution of Observed to Expected Ratios

Episode Type	Observed / Expected Ratio										
	Mean	Std. Dev.	Percentile								
			1st	5th	10th	25th	50th	75th	90th	95th	99th
All Final Episodes	1.00	0.52	0.38	0.49	0.55	0.66	0.85	1.18	1.67	2.02	2.89
Episodes with downstream acute (re)admission	1.58	0.67	0.71	0.85	0.94	1.12	1.41	1.85	2.42	2.86	3.93
Episodes without downstream acute (re)admission	0.91	0.42	0.38	0.48	0.53	0.64	0.79	1.03	1.45	1.76	2.45
Episodes with Post-Acute Care (IRF, LTCH, HH, SN)	1.20	0.56	0.44	0.56	0.64	0.81	1.06	1.45	1.92	2.26	3.10
Episodes without Post-Acute Care (IRF, LTCH, HH, SN)	0.80	0.38	0.36	0.45	0.50	0.60	0.71	0.87	1.13	1.45	2.40

The clinical themes analysis indicated that there is only a weak correlation between the index admission itself (correlation: 0.08) with the risk-adjusted cost. The clinical themes analysis also demonstrated that there is a strong correlation between the readmission (correlation: 0.47) and risk-adjusted cost. Correlation between Outpatient E&M services, procedures, and therapy (correlation: 0.26) and risk-adjusted cost is moderate. Finally, correlation between PAC: SNF (correlation: 0.34) and risk-adjusted cost is quite high, while the correlation between another PAC setting, home health (correlation: -0.18), is negative. The clinical theme analysis between PAC IRF/LTCH (correlation: 0.15) and risk-adjusted cost is moderate to low.

3.3.4 Interpretation

As expected, the average O/E cost ratio for episodes with downstream acute readmissions is higher than for episodes without downstream acute readmissions. This result demonstrates that the MSPB clinician measure is able to capture accurately higher resource use related to readmissions. Similarly, episodes with PAC services (i.e. HH, SNF, IRF, or LTCH) also have a substantially higher average ratio of observed to expected cost than episodes without PAC services.

The clinical themes analysis demonstrates that high risk-adjusted cost is strongly associated with themes related to readmissions and some types of post-acute care services (SNF), and linked – though more weakly, as expected -- to themes relating to services performed during the episode window and other types of post-acute care services (IRF/LTCH). Below are some more detailed interpretations of the correlations between clinical themes and risk-adjusted cost.

- Since the risk adjustment model adjusts for MS-DRGs for index admissions, the correlation between acute IP services performed including and during the index admission and the risk-adjusted cost is quite small.

- The correlation between Outpatient E&M services, procedures, and therapy is much higher because larger number of services provided during the episode window inevitably result in larger observed and expected costs. The costs for those services are not accounted for in the risk adjustment model since they are under direct influence of the clinician/clinician group practice providing care to the beneficiary.
- As expected, episodes that have readmissions will have high correlation with risk-adjusted cost. There are several reasons for this trend: (1) hospital readmissions are associated with higher costs and lead to larger observed over expected cost ratio, and (2) beneficiary's readmission is usually an indicator of poor health management from the attributed clinician/clinician group practice and therefore, is not adjusted for by the risk adjustment model.
- Between two types of post-acute care services, SNF and home health, SNF is associated with higher cost than home health services. Providers who make a decision to refer a beneficiary to SNF after hospitalization would incur much higher episode costs even after risk adjustment, since SNF is a less cost-efficient choice of PAC. Similarly, the correlation results show that sending a beneficiary to home health is inversely related to the risk-adjusted cost, since it is a more cost efficient service.

The results described above indicate that the measure may penalize clinicians who have higher rates of readmissions, while not disincentivizing the provision of appropriate care during the episode window or specifically during the index admission of an episode.

3.4 Exclusions Analysis

3.4.1 Method of Testing Exclusions

Exclusions are used in the MSPB clinician measure to ensure a homogenous patient population within the scope of the measure focus on inpatient hospitalizations/admissions and that episodes provide meaningful information to attributed clinicians or as part of data processing, to ensure that sufficient data are available to accurately determine episode spending and calculate risk adjustment for each episode. For the exclusions analysis, we focused on exclusions added to ensure a homogenous patient population. These exclusions, along with their rationales, are listed below:

- *Episodes where beneficiary death date occurred before the episode end.*
 - These episodes are excluded for all measures due to the potential to reflect inaccurately a clinician's performance. Episodes where the beneficiary died may be unusually high-cost, due to perimortem treatment costs, or unusually low-cost, due to the truncated episode window. Neither of these cases accurately reflects the efficiency of the clinician performing the treatment.
- *Episodes where the trigger claim is for an index admission that was not performed in an IP facility located in an eligible region*
 - Episodes with admissions in hospitals located outside the 50 states are excluded to remove episodes in which the beneficiary may have less access to care or receive care not reimbursed under the Medicare Part A and B.
- *Episodes in which Inpatient stay occurred in a non-Acute Hospital or in a Critical Access Care (CAH) hospital*
 - Episodes where the beneficiary's hospitalization did not occur in an acute care hospital facility are excluded from the measure calculation.

Given the rationales for these exclusions, we would expect these excluded episodes to have a different risk profile than the included episodes, such as a substantially higher or lower mean

cost, or a different distribution of costs (e.g., a long tail of high-cost episodes). For the exclusions, we examined the number of episodes and beneficiaries affected, as well as the distributions of observed cost and ratio of observed over expected spending (calculated by applying existing risk factor coefficients to the excluded episodes) for excluded episodes. We then compared the cost characteristics of the excluded episodes to those of episodes remaining after the described exclusions to assess the distinctness between the two patient cohorts.

3.4.2 Statistical Results from Testing Exclusions

Table 4 below presents observed cost statistics and O/E cost ratios for the MSPB clinician measure exclusions. Cost statistics are also provided for the remaining set of episodes after the described exclusions are applied for comparison.

Table 4: Cost Statistics for Measure Exclusions

Exclusion	Episodes		Observed Cost			O/E		
	#	%	Mean	Percentile		Mean	Percentile	
				10 th	90 th		10 th	90 th
All Episodes Meeting Triggering Logic	10,894,780	100.00%	\$21,248	\$6,414	\$40,329	1.04	0.51	1.78
Episodes in which Inpatient Stay had Transfers or Death Discharge Status Codes or episodes that overlapped with an IP Stay with Transfer or Death Discharge Status Codes	767,602	7.05%	\$31,459	\$9,094	\$62,750	1.24	0.45	2.46
Episodes in which beneficiary Death occurred within 30 Days Post Discharge	1,001,518	9.19%	\$23,509	\$8,080	\$44,537	0.88	0.41	1.59
Episodes in which Inpatient stay occurred in a non-Acute Hospital or in a Critical Access Care (CAH) hospital	1,252,558	11.50%	\$26,484	\$6,389	\$50,748	1.40	0.51	2.51
Episodes with Inpatient Facility located in Excluded Regions	31,496	0.29%	\$14,462	\$4,507	\$30,311	0.80	0.43	1.31
Remaining Episodes	7,012,126	64.36%	\$20,146	\$6,447	\$37,640	1.00	0.53	1.66

3.4.3 Interpretation

The statistical results indicate that the excluded episodes have different mean observed episode costs, which may be due to variation in payment that is not under the influence of the attributed clinician. These episodes can also be excluded due to clinical considerations to ensure a comparable patient cohort that will yield meaningful information to attributed clinicians. Further discussion of the results for each exclusion is provided below.

Episodes ending in death or transfer: The results indicate a large difference between the mean observed episode cost for death and non-death episodes: \$31,459 and \$23,509 for episodes related to beneficiaries who died or transferred compared to \$20,146 for the remaining set of episodes. Furthermore, the ratio of observed to expected episode cost for episodes with index admissions with discharges related to transfers or death is 1.24 indicating that the episodes are more costly than is predicted based on the risk adjustment model's consideration for patient characteristics and condition observed prior to the admission. For episodes in which the beneficiary passes within 30 days after discharge, the mean O/E cost ratio predicts the

episodes as being more expensive based on the patient's characteristics prior to the episodes start than observed. Including episodes with admissions, ending in death or transfer in the measure calculation may distort measure scores where patients require more resource use than can be predicted based on the patient case-mix upon admission and would give the appearance of less cost effective care. Inversely, when death occurs after discharge, including these episodes may skew a provider performance to look more efficient.

Episodes in which Inpatient stay occurred in a non-Acute Hospital or in a Critical Access Care (CAH) hospital: The mean observed cost of these episodes is \$6,338 more than for the remaining set of episodes, as expected. Since the risk adjustment model does not account for this difference in cost related to facility type, these cases are not included in the measure to ensure that only costs incurred in similar facilities are considered.

Episodes with Inpatient Facility located in excluded regions: Episodes with inpatient facilities in the following regions are excluded from the measure calculation because hospitals in those areas are not paid under the inpatient prospective payment system (IPPS): Puerto Rico, Virgin Islands, Canada, Mexico, Samoa, Guam, Marinas, and foreign countries.

3.5 Risk Adjustment or Stratification

3.5.1 Method of Controlling for Differences

Differences in case mix are controlled for using a statistical risk model with 110 risk factors

The risk adjustment model for the MSPB clinician measure broadly follows the CMS-HCC risk adjustment methodology, which is derived from Medicare Parts A and B claims and is used in the Medicare Advantage (MA) program. Although the MA risk adjustment model includes 24 age/sex variables, this risk adjustment model does not adjust for sex and so only includes 12 age categorical variables. Severity of illness is measured using HCCs, indicators of enrollment and long-term care status, and disease interactions.

The model includes 79 HCC indicators derived from the beneficiary's Parts A and B claims during the period 90 days prior to the episode start date and are specified in the CMS-HCC Version 22 (V22) 2016 model. Episodes for beneficiaries without a full 90-day lookback period are excluded from the measure. This 90-day period is used to measure beneficiary health status and ensures that each beneficiary's claims record contains sufficient data for risk adjustment purposes.

In addition, the risk adjustment model includes status indicator variables for whether the beneficiary qualifies for Medicare through Disability or has ESRD. The model also includes an indicator of whether the beneficiary recently required long-term care, defined as 90 days in a long-term care facility without being discharged to community for 14 days. Beneficiaries who need to reside in long-term care facilities typically require more intensive care than beneficiaries who live in the community. These enrollment and long-term care status variables are non-diagnostic based indicators of severity of illness.

The model also accounts for disease interactions between HCCs and/or enrollment status variables included in the MA model. These interactions are included because certain combinations of comorbidities increase costs more than is predicted by the HCC indicators alone.

As with the CMS-HCC model, the risk adjustment approach for this measure uses an ordinary least squares linear regression model. The predicted, or expected, cost is winsorized at 0.5th percentile to make sure episodes with unusually small predicted cost, which would lead to abnormally large O/E ratios, do not dominate certain clinicians' final score. The winsorized expected costs are renormalized to ensure the average expected episode cost is the same

before and after winsorizing. Then, extremely low- or high-cost outlier episodes with residuals below the 1st percentile or above the 99th percentile are excluded to reduce the effect of these episodes that deviate the most from their expected values in absolute terms. The expected cost after excluding these outliers is again renormalized to ensure that average expected costs are the same after outlier removal.

Finally, the risk adjustment model outlined above is performed separately for episodes within each MDC determined by the MS-DRG of the index admission.

Full details of the risk adjustment model are in the Measure Codes List File.¹² The National Summary Data Report (NSDR) Addendum includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model.¹³

3.5.2 Conceptual, Clinical, and Statistical Methods

We selected the CMS-HCC model based on previous studies evaluating its appropriateness for use in risk adjusting Medicare claims data. This model was developed specifically for use in the Medicare population, meaning that it accounts for conditions found in the Medicare population and is calibrated on Medicare fee-for-service beneficiaries. In addition, the CMS-HCC model is routinely updated for changes in coding practices (e.g., the transition from ICD-9 to ICD-10 codes) and is exhaustive on these code sets. Because the CMS-HCC model has already been extensively tested, we focus our testing on how the CMS-HCC model was adapted to the MSPB clinician measure methodology.

3.5.3 Conceptual Model of Impact of Social Risks

Our conceptual model of the impact of social risk factors is informed by both published, peer-reviewed literature and data analysis.

3.5.4 Statistical Results

The literature has extensively tested the use of the HCC model as applied to Medicare claims data. Although the variables in the HCC model were chosen to predict annual cost, CMS has also used this risk adjustment model in a number of other settings (e.g., ACOs, previous physician QRUR programs, and other measures such as NQF #2158: MSPB-Hospital cost measure). Recalling that the risk model relies on the existing CMS-HCC model, testing results for factors included in the CMS-HCC V22 2016 model can be found in the Pope et al (2011) report and the December 2018 CMS Report to Congress on risk adjustment in Medicare Advantage.^{14,15}

The results of the statistical analysis used to characterize our risk adjustment model can be found in the NSDR Addendum, which includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model.

¹² CMS, "Medicare Spending Per Beneficiary clinician Measure Codes List," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-mspb-clinician-codes-list.zip>.

¹³ CMS, "National Summary Data Report Addendum: 11 Episode-Based Cost Measures and Revised MSPB Clinician Measure," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html>.

¹⁴ Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

¹⁵ "Report to Congress: Risk Adjustment in Medicare Advantage", CMS <https://www.cms.gov/Medicare/Health-Plans/MedicareAdvgtgSpecRateStats/Downloads/RTC-Dec2018.pdf>.

3.5.5 Analyses and Interpretation in Selection of Social Risk Factors

Acumen analyzed gender, dual status, income, education, and unemployment as social risk factors (more information on these variables can be found in Section 3.1.8). Beneficiary gender and dual status were obtained from the EDB and CME. Information on income, education, and unemployment was obtained from ACS data and linked to episodes by census block group where possible to provide a more granular level of analysis than ZIP code.

The percentage of female beneficiaries range from 28.4 percent to 64.1 percent across the 23 of the 26 MDCs in this measure that reasonably occur for both genders (MDC 13 and MDC 14 are nearly 100 percent female as they are related to pregnancy, childbirth, and the female reproductive system, while MDC 12 is 0 percent female as it is related to the male reproductive system). For 22 out of 26 MDCs, the majority of the beneficiaries (57.38% - 84.91%) have non-dual status. The MDCs with a minority of non-dual status beneficiaries include MDC 14 – Pregnancy, Childbirth, and the Puerperium (9.6%), MDC 25 – Human Immunodeficiency Virus Infections (25.5%), MDC 19 – Mental Diseases and Disorders (37.5%), and MDC – Alcohol/Drug Use and Alcohol/Drug Induced Organic Mental Disorders (45.8%). Income level is categorized into high, medium, and low from the continuous average income variable in ACS; therefore, each category has 33.33 percent of observations. While 2.2 to 8.7 percent of beneficiaries across all MDCs are classified as having below a high school education level, between 91.3 and 97.8 percent of beneficiaries across MDCs are classified at a high school level or greater. Finally, 21.6 to 45.1 percent of beneficiaries have high unemployment designation (>10%).

Acumen examined the impact of including social risk factors into our risk adjustment model by running goodness of fit tests when different risk factors are added and compared to the base risk adjustment model, where the base risk adjustment model refers to the full standard set of risk adjustment variables from the CMS-HCC V22 2016 model, disability status, ESRD status, interaction variables, and recent long-term care use. Acumen ran a step-wise regression to include gender, dual status, gender + dual status, and gender + dual + income + education + unemployment + race, on top of the adapted CMS-HCC model. The step-wise regressions help evaluate individual as well as joint significance of the social risk factors. We examined the impact of including social risk factors into our risk adjustment model with T-test of individual significance and F-test of joint significance.

First, we analyzed the model coefficients and p-values for each of the base and social risk factor models to understand whether any of the social risk factor covariates are predictive of episode cost. The T-test and F-test revealed many significant p-values, indicating that social risk factors are likely predictive factors for determining resource use among beneficiaries for the relevant characteristic. However, the analysis also shows that the directions of the effects of social risk factors are not consistent. For example, high income episodes and high unemployment may display both significant positive and negative coefficients of spending across MDCs.

Secondly, we analyzed the impact of adding social risk variables on overall model performance by looking at the differences in the O/E cost ratio with and without social factors in the risk adjustment model. When including social risk factors in our risk adjustment regression, the minor differences in the O/E ratios, even for providers at high or low extremes of risk, indicates that social risk factor effects on the model performance are likely captured through existing risk adjustment variables. When including the social risk factors in risk adjustment, the ratio of observed over expected costs for 90.2 percent of TINs and 92.2 percent of TIN-NPIs changed by ± 0.01 or less. At a higher threshold, 99.6 percent of TINs and 99.7 percent of TIN-NPIs changed by ± 0.03 or less.

Finally, we analyzed the correlation between measure scores calculated with and without the social risk factors. The measure scores calculated with and without these social factors were highly correlated at both the TIN level (Spearman correlation coefficient of 0.997), and the TIN-NPI level (Spearman correlation coefficient of 0.998). These results indicate that the inclusion of social risk factors in the current risk adjustment model would have a limited effect on measure scores.

Due to the inconsistent direction and limited impact of social risk factor effects under the current risk adjustment model, we believe the MSPB clinician measure risk adjustment model sufficiently accounts for the effects of social risk factor on clinician measure scores.

3.5.6 Method for Statistical Model or Stratification Development

To analyze the validity of the current risk adjustment model, we examined three analyses: (1) R-squared and adjusted R-squared for the regression models, (2) predictive ratios and O/E cost ratios to examine the fit of the models at different levels of patient complexity, and (3) coefficient estimates, standard errors, and p-values for each MDC.

- 1) *R-squared and adjusted R-squared* were calculated for each MDC. The results should be evaluated in the context of the service exclusion rules for each MDC, which indicate which costs are counted in the measures and which costs are not counted. This is an important distinction from all-cost measures, as a low R-squared does not necessarily indicate that a measure reflects variation unrelated to clinical care, while a high R-squared does not necessarily indicate the opposite; instead, the risk adjustment models must be evaluated in concert with the service exclusion rules. These results are provided in Section 3.5.7.
- 2) *Predictive ratios and O/E cost ratios* were calculated for each “risk decile” for the episode group. A “risk decile” is based on the risk scores, which indicate how costly episodes are expected to be, as predicted through risk adjustment. After arranging episodes into deciles based on their risk score, we calculated the predictive ratios and average O/E cost ratios for each decile. The predictive ratio aims to examine the fit of the model at different levels of patient complexity to examine the model’s ability to predict both very low and high cost episodes, and is calculated using the formula of average (expected cost)/average (observed cost) for all episodes in each decile. Similarly, the O/E cost ratio demonstrates the model’s prediction accuracy, and is calculated using the formula of average (observed cost/expected cost) for all episodes in each decile. These are discussed in Sections 3.5.8 and 3.5.9.
- 3) *Coefficient estimates, standard errors, and p-values* were run for each MDC to consider the extent to which the coefficients for the risk factors are predictive of episode cost. Results for individual risk adjustment variables should be viewed in the context of the entire model and set of MDCs, rather than being analyzed individually. For instance, coefficients indicate the incremental effect of a model variable, holding all other variables fixed. As another example, interactions between model variables must be interpreted in concert with the effects of those variables in isolation.

The results of these analyses are presented in the NSDR Addendum to aid in the overall assessment of the predictive ability of the risk adjustment models.

3.5.7 Statistical Risk Model Discrimination Statistics

The range of R-squared values for the MSPB clinician cost measure risk adjust models, calculated by dividing explained sum of squares by total sum of squares ranges between 0.11 – 0.57 across the MDCs. The adjusted R-squared range is 0.11 – 0.57.

The NSDR Addendum also includes regression coefficients and standard errors for each of the covariates used in the risk adjustment model. More information on discrimination testing for the CMS-HCC model can be found at Pope et al. 2011.¹⁶

3.5.8 Statistical Risk Model Calibration Statistics

We interpret calibration as how accurately the risk model's predictions match the actual episode cost. We calculate the average O/E cost ratio for each risk decile to demonstrate the model's prediction accuracy. The average observed to expected cost is generally close to one, 0.99 to 1.01, across risk factors, indicating that the model is accurately predicting actual episode cost for that risk factor. Full results can be seen the NSDR Addendum.

3.5.9 Statistical Risk Model Calibration – Risk Decile

Analysis of predictive ratios by risk decile for the measure shows that the model has consistent predictive ratios across risk score deciles, with each decile having a predictive ratio between 0.99 and 1.01.

Table 5: Distribution of the MSPB Clinician Measure Scores

Level	Provider Count	Mean Score	Score Percentile						
			1st	10th	25th	50th	75th	90th	99th
TIN	20,853	\$18,838	\$14,942	\$16,798	\$17,730	\$18,701	\$19,791	\$20,982	\$23,939
TIN-NPI	127,529	\$19,365	\$15,430	\$17,149	\$18,150	\$19,262	\$20,446	\$21,659	\$24,420

3.5.10 Interpretation

The R-squared values for the model, which measure the percentage of variation in results predicted by the model, are higher than the values presented in similar analyses of risk adjustment models. As noted in Section 3.5.6, these results should be interpreted alongside service exclusion rules, which remove clinically unrelated services, so the resulting variation is reflective of variation related to factors within a clinician's reasonable influence.

As demonstrated in Section 3.5.8 and 3.5.9, the average O/E cost ratios and the predictive ratios for all risk deciles are close to one. Predictive ratios close to one indicate that expected spending is accurately predicting observed spending. Overall, the results show that the model is accurately predicting spending, regardless of overall risk level.

3.6 Identification of Meaningful Differences in Performance

3.6.1 Method

Our method of determining clinically meaningful differences in the MSPB clinician measure scores consists of stratifying the measure scores by meaningful characteristics and investigating the clinician score distribution by percentile. Stratification is performed for each of the following characteristics: urban/rural, census division, census region, risk score, and the number of episodes attributed to the clinician. We analyze the distribution of measure scores for clinicians defined by these characteristics, as well as for the overall measure.

The purpose of this analysis is to ensure that there is a sufficiently large difference in measure scores among clinicians to determine a meaningful difference in performance. In addition, this

¹⁶ Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

analysis looks to confirm that the measure behaves as expected with respect to meaningful clinician characteristics.

3.6.2 Statistical Results

Key findings show that, generally, there is a large performance difference among clinicians in the MSPB clinician measure:

- (i) the 99th percentile of the measure score is nearly 1.6 times the 1st percentile at both the TIN level and TIN-NPI levels; and
- (ii) the 90th percentile of the MSPB clinician measure score is approximately 1.26 times the 10th percentile at both the TIN and TIN-NPI level.

These results indicate there is large potential for saving Medicare spending.

The results also show that there is not systemic regional difference in clinician score. For instance, the difference in the mean scores for clinicians across nine census divisions (excluding 'Unknown') are within less than \$1,290 for both the TIN and TIN-NPI testing (i.e., \$17,992 - \$19,281 at the TIN level and \$18,933 - \$19,895 at the TIN-NPI level). Similarly, clinicians in urban areas seem to perform comparably to those in rural areas with less than a \$580 difference in mean score.

In terms of other clinician characteristics, analysis of clinicians by number of episodes indicates that clinicians with more episodes perform similarly to those who have fewer episode. We also analyzed clinicians by risk score decile, as variation by risk score decile could indicate that the risk adjustment model is over- or under-correcting for clinicians with systematically riskier patients. Measure scores also show little variation by risk score decile, with a range in mean TIN score of \$17,765 to \$19,362 and a range in mean TIN-NPI score of \$18,030 to \$19,949, indicating that the risk adjustment model is overall functioning as intended. Full results can be seen in the NSDR.¹⁷

3.6.3 Interpretation

There is clinically and practically significant variation in MSPB clinician measure scores, indicating the measure's ability to capture differences in performance. Our findings regarding variation in measure scores are consistent with expert clinician input. The results indicate that clinicians are not being penalized or rewarded due to risk score decile given the current MSPB clinician measure design. For example, the measure is not systematically discriminating against certain types of clinicians (clinicians practicing in rural vs. urban setting, or small vs. large providers). At the same time, the measure does capture meaningful differences in resource use and, thus, provides actionable feedback to clinicians on how to improve their performance on the measure through care practice improvements.

3.7 Missing Data Analysis and Minimizing Bias

3.7.1 Method

Since CMS uses Medicare claims data to calculate the MSPB clinician measure, Acumen expects a high degree of data completeness. To ensure further that we have complete and accurate data for each beneficiary who opens an episode, Acumen excludes episodes where

¹⁷ CMS, "National Summary Data Report: 11 Episode-Based Cost Measures and Two Revised Cost Measures, Updated Following Field Testing (Oct-Nov 2018)," *MACRA Feedback Page*, <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/value-based-programs/macra-mips-and-apms/macra-feedback.html>.

beneficiary date of birth information (an input to the risk adjustment model) cannot be found in the EDB or the beneficiary death date occurs before the episode trigger date.

The MSPB clinician measure also excludes episodes where the beneficiary is enrolled in Medicare Part C or has a primary payer other than Medicare in the 90-day lookback period and episode window. In such situations, Medicare Parts A and B claims data may not capture the complete clinical profile for the beneficiary needed to capture the clinical risk of the beneficiary in risk adjustment. Furthermore, Parts A and B claims data may not capture all Medicare resource use if some portion of the beneficiary's care is covered under Medicare Part C.

3.7.2 Missing Data Analysis

The table below presents the frequency of missing data across the four categories of missing data, which caused episodes to be excluded from the MSPB clinician measure. Frequency is presented in terms of the number of episodes excluded due to missing data, as well as the number of TINs and TIN-NPIs who had at least one episode excluded due to missing data. The missing data categories are:

- Beneficiary date of birth is missing
- Beneficiary death date occurred before the trigger date
- Beneficiary has a primary payer other than Medicare during the episode window or in the 90-day lookback period
- Beneficiary was not enrolled in Medicare Parts A and B, or was enrolled in Part C, during the 90-day lookback period and episode window

Table 6: Missing Data Categories for the MSPB Clinician Measure

Exclusion	# Episodes	# TINs	# TIN-NPIs
Missing birth date	*	*	*
Death before trigger	84	40	96
Primary payer other than Medicare	1,150,251	42,865	301,702
Not continuously enrolled in Parts A and B	1,534,091	43,004	319,987

*denotes that there were fewer than 11 episodes

3.7.3 Interpretation

As the MSPB clinician measure is calculated with Medicare claims data, Acumen expects a high degree of data completeness, which is supported by the limited frequency of missing data for birth date and invalid beneficiary death date information above. Additionally, the measure will remove beneficiaries that may have gaps in the Medicare claims history due to alternate enrollment. Acumen takes measures to address cases of missing or inaccurate information in claims data.

4.0 Feasibility

4.1 Data Elements Generated as Byproduct of Care Processes

The data elements used in this measure are generated, collected, and/or used by healthcare personnel during the provision of care (e.g., blood pressure, laboratory values, diagnosis, depression score). The data collected during care provision are then translated into the appropriate coding system (e.g. ICD-10 diagnoses, MS-DRGs) for use in Medicare claims.

4.2 Electronic Sources

All data elements are in defined fields in electronic claims.

4.3 Data Collection Strategy

4.3.1 Data Collection Strategy Difficulties

Lessons and associated modifications may be categorized into three types: data collection procedures, handling of missing data, and sampling data associated with beneficiaries who died during an episode of care.

4.3.1.1 Data Collection

Acumen receives claims data directly from the Common Working File (CWF) maintained at the CMS Baltimore Data Center. Medicare claims are submitted by healthcare providers to a Medicare Administrative Contractor (MAC), and are subsequently added to the CWF. However, these claims may be denied or disputed by the MAC, leading to changes to historical CWF data. In rare circumstances, finalizing claims may take many months, or even years. As a result, it is not practical to wait until all claims for a given month are finalized before calculating this measure. As such, there is a trade-off between efficiency (accessing the data in a timely manner) and accuracy (waiting until most claims are finalized) when determining the length of the time (i.e., the “claims run-out” period) after which to pull claims data. To determine the appropriate claims run-out period, Acumen has performed testing on the delay between claim service dates and claims data finalization. Based on this analysis, Acumen uses a run-out period of three months after the end of the calendar year to collect data for development and testing purposes. If this measure were used in a CMS program, calculation and reporting would be done in line with that program’s reporting practices.

4.3.1.2 Missing Data

This measure requires complete beneficiary information, and a small number of episodes with missing data are excluded to ensure completeness of data and accurate comparability across episodes. For example, episodes where the beneficiary was not enrolled in Medicare Parts A and B for the 90 days prior to the episode start date are not included in this measure. This enables the risk adjustment model to adjust accurately for the beneficiary’s comorbidities using data from the previous 90 days of Medicare claims. Additionally, the risk adjustment model includes a categorical variable for beneficiary age bracket, so episodes for which the beneficiary’s date of birth cannot be located are not included in this measure.

4.3.1.3 Sampling

During measure testing, Acumen noted that episodes in which the beneficiary died prior to the episode end date exhibited different cost distributions compared to other episodes. To avoid this effect’s potential impact on clinician scores, this measure does not include episodes for which the beneficiary’s date of death occurs prior to the end of the episode window.

5.0 Usability and Use

5.1 Use

5.1.1 Current and Planned Use

The measure was developed for potential use in the Merit-based Incentive Program (MIPS), under a contract with CMS.

5.1.2 Feedback on the Measure and Development Process

5.1.2.1 Technical Assistance Provided During Development or Implementation

Development: Field Testing

Acumen and CMS conducted a national field test of 11 episode-based cost measures and two population-level cost measures, including the Total Per Capita Cost (TPCC) measure and the MSPB clinician measure, developed during 2018 for a 35-day comment period (October 3 to November 5, 2018). We provided MSPB clinician Field Test Reports to a sample of eligible clinician groups and clinicians.¹⁸ Each report included information for the MSPB clinician measure if the clinician or clinician group was attributed 35 or more episodes. The testing sample was selected to balance coverage and reliability, since a key goal of field testing was to test the measures with as many stakeholders as possible. This sampling technique was used for field testing only and does not determine case minimums used for any potential program implementation. The number of field test reports shared with the public was:

- MSPB Clinician: 148,382; 20,852 TINs; 127,530 TIN-NPIs

All stakeholders, including those who did not receive a field test report, could review a mock field test report that was posted on the CMS website. Other public documentation posted during field testing included: measure specifications (comprising a Draft Cost Measure Methodology document and a Draft Measure Codes List file), a Frequently Asked Questions document, and a Fact Sheet.¹⁹ During field testing, Acumen conducted education and outreach activities including a national webinar, office hours with specialty societies, and Help Desk support.

5.1.2.2 Technical Assistance with Results

Field Testing

During the feedback period, 148,381 field test reports for the MSPB clinician measure were downloaded by 436 clinician groups (TINs) and 4,717 clinicians (TIN-NPIs). Stakeholder comments from field testing were summarized for the workgroup to consider in recommending refinements to the measures based on the testing data and feedback.

The following sections offer more details on the contents of each report and describe the education and outreach efforts associated with the field testing feedback period.

Data Provided During Field Testing

Each Field Test Report contained the following information:

- Results on clinician/clinician group measure scores
- Clinician Cost Breakdown by Claim Type
- Clinician Cost Breakdown by MDC

¹⁸ The field test reports were available for download from the CMS Enterprise Portal: <https://portal.cms.gov/wps/portal/unauthportal/home/>.

¹⁹ The Measure Development Process, Frequently Asked Questions, and Fact Sheet documents are posted on the MACRA Feedback Page: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-Feedback.html>.

- Clinician Cost Breakdown by Categories of Service
- National Distribution of MSPB Clinician Measure Scores

A mock field test report can be viewed on the CMS MACRA Feedback webpage.²⁰

Education and Outreach

Acumen directly conducted outreach via email to tens of thousands of stakeholders using the stakeholder contact list developed through previous education and outreach and clinician engagement efforts, as well as CMS, Quality Payment Program, and other available listservs. More detail on this outreach can be found in the Field Test Summary Report on the CMS MACRA Feedback webpage.

Acumen and CMS hosted two office hour sessions in October 2018, to provide an overview of field testing to specialty societies, discuss what information their members would be particularly interested in, and answer any questions. Across all both office hours sessions, there were 50 attendees.

Acumen and CMS hosted a national field testing webinar on October 9, 2018 to provide an overview of the measures being field tested and the information available for public comment. The webinar consisted of an hour-long presentation, outlining (i) the cost measure development activities, (ii) field testing activities, (iii) how to access and understand the confidential field test reports, and (iv) the contents of the reports. The presentation was followed by a 30-minute Q&A session. Around 85 comments and questions were received via webinar chat and on the phone.

A post-field testing webinar was held on March 27, 2019 to provide an update on the measures following field testing. The webinar consisted of approximately 60 minutes to provide an overview of the basics of measure construction, highlight refinements made after field testing, and provide a summary of testing done on the measures. The presentation was followed by a 30-minute Q&A portion.²¹

5.1.2.3 Feedback on Measure Performance and Implementation

Field Testing

In total, Acumen received 67 survey responses and 25 comment letters, including many from specialty societies representing large numbers of potentially attributed clinicians.

Survey responses and comment letters were collected via an online survey, which contained general and detailed questions on the reports themselves, questions on the supplemental documentation, and questions on the measure specifications.

Pre-Rulemaking

CMS received 14 comments specifically on the MSPB clinician cost measure included in the Measures Under Consideration List released in December 2018. After the MAP Clinician Workgroup meeting in December 2018, there was another public comment period on their preliminary recommendations, which received 6 comments specific to the MSPB clinician cost measure.²² Stakeholders were able to submit their comments via the NQF website.

²⁰ CMS, “Revised MSPB Clinician Measure Mock Field Test Report,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/Mock-report-for-revised-MSPB-Clinician.pdf>

²¹ CMS, Webinar Recordings, Slides, and Transcripts, *QPP Webinar Library* <https://qpp.cms.gov/about/webinars>.

²² Measure Applications Partnership, *National Quality Forum*. https://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx.

5.1.2.4 Feedback from Providers being Measured

Field Testing

The Field Testing Feedback Summary Report presents all feedback gathered during the field testing period. The following list synthesizes some of the key points that were raised through the field testing feedback period:

- *Field test reports and supplementary materials present useful information for understanding the measure, though reduced complexity could encourage more clinician participation.* Stakeholders praised the content of the field test reports and supplementary materials and commented that overall the methodological changes are described in a clear and concise manner. However, the complexity of the information presented in the reports was a challenge for some stakeholders. Specifically, several stakeholders noted that the methodology to attribute MSPB clinician episodes at the TIN and/or TIN-NPI level was too complex. Some commenters also requested additional information to increase their understanding of the measure.
- *Service exclusion codes and logic developed for the MSPB clinician episodes make the measure more actionable.* Stakeholders expressed appreciation for the development of codes and logic that define a list of services that are unlikely to be influenced by the clinician's care decisions and exclude clinically unrelated services to calculate episode observed cost. In addition to this comment, one stakeholder noted some additional codes for removal from the measure.
- *Improved measure better captures clinicians responsible for a beneficiary healthcare cost, but can be further refined.* One commenter expressed concern that the measure might not reflect the fact that the clinician group practice keeps moderately sick patients out of the inpatient setting and has a disproportionate number of sicker people with inpatient stays. The commenter also indicated that under the current measure calculation methodology this would make the group practice have an unreasonably high MSPB clinician measure score.

5.1.2.5 Feedback from Other Users

Pre-Rulemaking

The MAP recognized the importance of cost measures to the MIPS program and conditionally supported the MSPB clinician cost measure pending NQF endorsement. Specifically, the MAP urged CMS to continue testing the changes to this measure, which are removing costs that are unlikely related to the clinician and a new attribution model, to ensure that they produce the intended results. In particular, MAP noted the need to ensure the measure demonstrates validity and reliability at the National Provider Identifier (NPI) level. MAP also noted the desire to avoid double counting clinician costs in the total cost measures and the episode-based cost measures and for CMS to consider consolidating the MSPB clinician measure with the Total Per Capita Cost measure also used in MIPS to avoid overlap. MAP suggested that CMS should monitor for unintended consequences to patients such as under treatment, impact on technology innovation, and access to treatment for high-risk, high-resource use patients. Lastly, MAP urged CMS to continuously test and refine the risk adjustment model and incorporate social risk factors, when appropriate.

5.1.2.6 Consideration of Feedback

Field Testing

After completing field testing, Acumen compiled the feedback provided through the survey and comment letters into a measure-specific report, which was then provided to the expert clinician workgroup, along with empirical analyses to inform their discussion and evaluation of any refinements needed to ensure that the measure is capturing what it was intended to capture.

After careful consideration of field testing analyses and stakeholder feedback, no refinements were made to the measure after field testing.

5.2 Usability

5.2.1 Improvement

This revised measure has not yet been implemented, and as such has not had influence over performance. The number of clinicians in the Quality Payment Program varies by performance period. As outlined in the 2017 Quality Payment Program Reporting Experience, there were 1,057,824 MIPS eligible clinicians receiving a MIPS payment adjustment in 2017.²³ This report refers to the version of TPCC currently in use in MIPS. As clinicians have choices on how to participate in the Quality Payment Program (e.g., through MIPS or the Advanced APMs, as groups or individuals), the exact number, and percentage of clinicians who will receive a performance score on this measure will only be confirmed after the end of each performance period.

5.2.2 Unexpected Findings

n/a. There were no unexpected findings during the development and testing of this measure

5.2.3 Unexpected Benefits

n/a. There were no unexpected benefits during the development and testing of this measure.

²³ CMS, “2017 Quality Payment Program Reporting Experience,” *QPP*, <https://qpp-cm-prod-content.s3.amazonaws.com/uploads/491/2017%20QPP%20Experience%20Report.pdf>

6.0 Related and Competing Measures

6.1 Relation to Other Cost Measures

There is currently a related MSPB – Hospital measure used to measure facilities resource use implemented in the Value-Based Purchasing (VBP) program. This measure is NQF endorsed (NQF #21580). The revised MSPB clinician measure has opportunity to work in concert with the MSPB hospital measure by having both the facility and the clinician's role and responsibility measured when managing the care of a beneficiary admitted into the inpatient setting.

6.2 Harmonization

The MSPB-Hospital and MSPB clinician measures are closely aligned. Both cost measures use the same time window (three days prior to the index admission to 30 days after discharge) during which cost is assessed in an episode. In addition, the risk adjustment models for the two measures are closely aligned, adjusting for nearly an identical set of risk factors using the same algorithm for HCC construction.

6.3 Competing Measures

n/a

Contact Information

Measure Steward Point of Contact

Organization: Centers for Medicare & Medicaid Services

Name: Joel Andress

Email Address: joel.andress@cms.hhs.gov

Phone Number: (410) 786-5237

Developer Point of Contact

Organization: Acumen, LLC

Email Address: macra-cost-measures-info@acumenllc.com

Phone Number: (650) 558-8882

Other Additional Information

Technical Expert Panel Members

Adolph Yates, American Academy of Orthopaedic Surgeons
Alan Lazaroff, American Geriatrics Society
Allison Madson, American Society of Cataract and Refractive Surgery
Alvia Siddiqi, American Academy of Family Physicians
Anupam Jena, Harvard Medical School
Caroll Koscheski, American College of Gastroenterology
Chandy Ellimoottil, American Urological Association
Diane Padden, American Association of Nurse Practitioners
Dyane Tower, American Podiatric Medical Association
Edison A. Machado, Jr., The American Health Quality Association
Jackson Williams, Dialysis Patient Citizens
James Naessens, Mayo Clinic
John Bulger, American Osteopathic Association
Juan Quintana, American Association of Nurse Anesthetists
Kata Kertesz, Center for Medicare Advocacy
Kathleen Blake, American Medical Association
Mary Fran Tracy, National Association of Clinical Nurse Specialists
Parag Parekh, American Society of Cataract and Refractive Surgery
Patrick Coll, University of Connecticut Health Center
Shelly Nash, Adventist Health System
Sophie Shen, Johnson and Johnson Health Care Systems, Inc.

MSBP Service Refinement Workgroup Members:

Adolph Yates, American Association of Hip and Knee Surgeons
Clemens Schirmer, American Association of Neurological Surgeons
Dheeraj Mahajan, AMDA – The Society for Post-Acute and Long-Term Care Medicine
Jennifer Bracey, Society of General Internal Medicine
Steve Sentovich, American Society of Colon and Rectal Surgeons
Marc Raphaelson, American Academy of Neurology
Evan Lipsitz, Society of Vascular Surgery
Caroll Koscheski, American College of Gastroenterology
Richard Dutton, American Society of Anesthesiologists
Kathleen Blake, American Medical Association
Anupam Jena, Harvard Medical School
Alec Koo, American Urological Association

Sanjay Samy, Society of Thoracic Surgeons
Juana Hutchinson-Colas, American Urogynecologic Society
Naakesh Dewan, American Psychiatric Association
Jayme Lieberman, American College of Surgeons
Tracey Weisberg, American Society of Clinical Oncology
William Borden, American College of Cardiology
Anders Chen, Society of General Internal Medicine
Nancy Greenwell, American Academy of Family Physicians
Stephen Lahey, Society of Thoracic Surgeons
Terry "Lee" Mills, American Academy of Family Physicians
Peter Ray, American Society of Plastic Surgeons
Barbara Spivak, American Medical Association
Robert Lorenz, American Academy of Otolaryngology – Head and Neck Surgery