

Total Per Capita Cost 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.....	7
3.0	Scientific Acceptability	8
3.1	Data Sample Description	8
3.1.1	Type of Data Used for Testing.....	8
3.1.2	Specific Dataset Used for Testing	8
3.1.3	Dates of the Data Used in Testing.....	8
3.1.4	Levels of Analysis Tested.....	8
3.1.5	Entities Included in the Testing and Analysis.....	8
3.1.6	Patient Cohort Included in the Testing and Analysis.....	8
3.1.7	Sample Differences	9
3.1.8	Social Risk Factors Included in Analysis	9
3.2	Reliability Testing.....	10
3.2.1	Level of Reliability Testing	10
3.2.2	Method of Reliability Testing	10
3.2.3	Statistical Results from Reliability Testing	11
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	13
3.3.4	Interpretation	14
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	15
3.5.1	Method of Controlling for Differences	15
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.....	17
3.5.6	Method for Statistical Model or Stratification Development.....	18
3.5.7	Statistical Risk Model Discrimination Statistics	18
3.5.8	Statistical Risk Model Calibration Statistics	18
3.5.9	Statistical Risk Model Calibration – Risk Decile.....	18
3.5.10	Interpretation.....	19
3.6	Identification of Meaningful Differences in Performance.....	19
3.6.1	Method	19

3.6.2 Statistical Results	19
3.6.3 Interpretation	20
3.7 Missing Data Analysis and Minimizing Bias.....	20
3.7.1 Method	20
3.7.2 Missing Data Analysis	20
3.7.3 Interpretation	20
4.0 Feasibility.....	22
4.1 Data Elements Generated as Byproduct of Care Processes	22
4.2 Electronic Sources	22
4.3 Data Collection Strategy	22
4.3.1 Data Collection Strategy Difficulties.....	22
5.0 Usability and Use	23
5.1 Use	23
5.1.1 Current and Planned Use	23
5.1.2 Feedback on the Measure and Development Process	23
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 Total Per Capita Cost (TPCC) 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

Total Per Capita Cost (TPCC) measure

1.3 Type of Measure

Cost/Resource Use

¹ CMS, “Total Per Capita Cost Measure Methodology,” *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-TPCC-measure-specs.zip>
CMS, “Total Per Capita Cost 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-TPCC-measure-specs.zip>.

2.0 Importance

2.1 Evidence to Support the Measure Focus

2.1.1 Measure Description

The revised TPCC measure evaluates overall cost of care delivered to a beneficiary with a focus on the primary care they receive from their provider(s). The TPCC measure was revised according to the CMS Blueprint requirements for comprehensive re-evaluation of measures.² The TPCC cost measure score is a clinician's risk-adjusted and specialty-adjusted monthly cost averaged across all of the beneficiary months attributed to the clinician. For this total cost measure, all of a beneficiary's costs are assigned to a clinician during the months in which they are responsible for the beneficiary's primary care. A specific set of clinical events that indicate the start of a primary care relationship between the beneficiary and clinician open or 'trigger' the risk window. The portion of the risk windows that overlap the measurement period is used to identify the beneficiary's months for which the clinician will be attributed and measured. Beneficiary populations eligible for the TPCC 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.³ 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 periods in which they can be considered responsible for a beneficiary, or if lower spending and better care quality can be delivered through changes in clinical practice.

Research shows that primary care management in certain settings, such as Patient-Centered Medical Homes (PCMH), has brought about measurable reductions to the total cost of care by reducing utilization of high-cost services and in some cases, by directing patients to lower cost hospitals.⁴ With this research-based evidence available for certain settings, a key question for policymakers is whether primary care management would achieve similar results across a wider variety of settings. In light of this question, a measure that captures the cost performance of primary care providers across a range of settings can help to confirm the benefits of effective primary care management. Given that, as noted above, clinicians are often unaware of how their choices affect the total costs of care, such a measure can help guide primary care providers towards practices that reduce costs, while maintaining or improving quality.

TPCC is a broad measure that focuses on measuring the performance of clinicians delivering primary care services, which can include both primary care and specialty clinicians. By allowing more clinicians to have their cost performance measured, this broad measure complements

² CMS, "Blueprint for the CMS Measures Management System," <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/Blueprint.pdf>.

³ Crosson, FJ. "Change the microenvironment. Delivery system reform essential to control costs." *Mod Healthc.*, vol. 39, no. 17, 2009, pp. 20-1

⁴ "Valuation of Care Management Performed by Primary Care Services: An Issue Brief." American Academy of Family Physicians, 2018.

more specific episode-based cost measures, which measure the performance of a subset of specialties concentrated around a specific condition or procedure. In complementing episode-based cost measures, all-cost measures, such as TPCC become an important means to enhance the coverage of patients and effectively incentivize improvements in the efficiency of care delivery in Medicare.

Another key opportunity presented by a cost performance measure for primary care is the opportunity to reward primary care providers for delivering value and to thereby improve patients' access to primary care services. As noted by MedPAC, beneficiaries experience more difficulty accessing primary care than with accessing specialty care.⁵ More specifically, 1.3 percent of the Medicare population reported a "big problem" finding a primary care doctor, while just 0.9 percent of this population reported such a problem in finding a specialist in 2017. Relatedly, among patients desiring to switch primary care providers, some patients felt that this was not an option due to long wait times or due to practices being closed to new patients. This may be related to another fact that MedPAC observes in the same report, which is that the Physician Fee Schedule's orientation to discrete services with a clear beginning and end does not support primary care, with its need for ongoing care coordination for a group of patients. Given this, MedPAC recommended the establishment of a per beneficiary payment for primary care practitioners to replace the expired Primary Care Incentive Payment (PCIP) program. This program provided a 10 percent bonus on fee schedule payments for some E&M services delivered by primary care practitioners. While the establishment of such a revised payment policy for primary care management might be an optimal solution to increase the availability of primary care, it may take substantial time to implement. Given this, it is particularly important to utilize an existing measure of the cost performance of primary care clinicians to identify and provide financial incentives for good performance.

2.2 Performance Gap

2.2.1 Rationale

Effective primary care management can support Medicare savings in a number of ways, including through improvements in the treatment of chronic conditions by obviating the need for high-cost hospital or emergency department services. More effective primary care management can also direct a greater proportion of patients to lower hospital costs for inpatient services. Given the potential for decreasing spending through improvements in primary care delivery, the TPCC measure allows for a savings opportunity by capturing the broader healthcare costs influenced by primary care.

A TPCC measure was originally used in the Physician Value-Based Payment Modifier (VM) Program and reported in the annual Quality and Resource Use Reports (QRURs). With the introduction of the Quality Payment Program, the current version of the TPCC measure was finalized with minor adaptations from the VM Program and added to MIPS. Subsequent rules have noted that the current TPCC would be refined based on stakeholder input. This MJF presents testing results for the revised TPCC measure, refined based on stakeholder input. A summary of the differences between the revised and the current TPCC measure can be found in Appendix A of the Cost Measure Methodology for the revised TPCC on the CMS MACRA Feedback webpage.⁶

⁵ "Report to the Congress: Medicare Payment Policy," *MedPAC*, 2018, http://www.medpac.gov/docs/default-source/reports/mar18_medpac_entirereport_sec.pdf

⁶ CMS, "Total Per Capita Cost Measure Methodology," *MACRA Feedback Page*, <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/2019-revised-TPCC-measure-specs.zip>.

2.2.2 Performance Scores

Performance scores are provided for 77,479 clinician group practices (identified by Tax Identification Number [TIN]) and 326,649 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 Merit-based Incentive Payment System (MIPS) eligible clinician specialty, and do not reflect other MIPS eligibility criteria (e.g., Advanced Alternative Payment Model participation). This table uses a testing volume threshold of 20 beneficiaries.

Table 1: Distribution of Performance Scores

Metric	TIN	TIN-NPI
Mean Score	\$1,053	\$1,110
Standard Deviation	\$247	\$291
Score IQR	\$240	\$281
Score Percentiles		
10 th	\$791	\$811
20 th	\$889	\$915
30 th	\$950	\$981
40 th	\$998	\$1,035
50 th	\$1,040	\$1,086
60 th	\$1,084	\$1,137
70 th	\$1,132	\$1,196
80 th	\$1,196	\$1,274
90 th	\$1,309	\$1,412

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 TPCC measure uses Medicare Part A and Part B claims data maintained by CMS. This claims data is used to attribute beneficiary months, calculate beneficiary's monthly costs, and construct risk adjusters. Data from the EDB are 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 models also account 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 identify beneficiaries that should be risk adjusted through the CMS-HCC institutional model.

For measure testing, data from the American Census, American Community Survey (ACS), and CME are used in analyses evaluating patient cohort.

3.1.3 Dates of the Data Used in Testing

The measurement period for the revised TPCC testing is October 1, 2016 through September 30, 2017

3.1.4 Levels of Analysis Tested

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

3.1.5 Entities Included in the Testing and Analysis

77,479 clinician group practices and 326,649 practitioners were included in the analyses. Clinicians and clinician groups were included in testing if they were attributed 20 or more TPCC beneficiaries during the measurement period. Beneficiaries from all 50 States and D.C. receiving evaluation and management care indicative of primary care were included, with their respective costs evaluated from all claim settings.

3.1.6 Patient Cohort Included in the Testing and Analysis

26,647,274 Medicare beneficiaries (with 305,869,065 beneficiary months) were included in TIN level testing and analysis, and 26,398,076 beneficiaries (with 298,109,075 beneficiary months) were included in TIN-NPI level measure testing.

The beneficiary population eligible for the TPCC attribution consists of Medicare beneficiaries enrolled in Medicare Parts A and B (but not Part C) receiving evaluation and management services that indicate a primary care relationship. Beneficiaries were included in the sample if they met a set of inclusion criteria (listed below) meant to ensure completeness of data.

The inclusion criteria are:

- The beneficiary has Medicare as their primary payer for the entire measurement period.
- The beneficiary was continuously enrolled in Medicare Parts A and B and any instance of partial enrollment was the result of either new enrollment or death only.

- The beneficiary date of birth is not missing.
- The beneficiary death date did not occur before the measurement period.
- The beneficiary did not reside outside the United States or its territories during any month of the measurement period
- The beneficiary is not covered by the Railroad Retirement Board

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

This analysis shows that the TPCC measure's inclusion criteria have only a minimal effect on the percentage of total beneficiaries of any particular demographic at the TIN level. To illustrate, the percentage of beneficiaries aged 65 to 69 without applying the inclusion criteria is 30.7 percent, compared to 31.0 percent with the inclusion criteria at TIN level testing. The breakdown of male and female beneficiaries remains within a 0.01 percentage point difference when comparing the application of inclusion criteria. The percentage of beneficiaries identified as female without applying the inclusion criteria is 56.3 percent, compared 56.4 percent with inclusion criteria at TIN level testing. These results indicate that there is minimal shift in patient characteristics as a result of using the inclusion criteria listed above at TIN level testing.

3.1.7 Sample Differences

n/a

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 percent with < high school education is the highest for a given Census Block Group
 - Education = High School: when percent with only high school is the highest
 - Education > High School: when percent with > high school is the highest
- Employment (ACS)
 - Unemployment Rate > 10 percent
 - Unemployment Rate <= 10 percent
- 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 TPCC 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.⁷ The FY 2018 Medicare FFS program proper payment rate was 91.9 percent.⁸ 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

⁷ Comprehensive Error Rate Testing 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>

⁸ Ibid.

$$\sigma_b^2$$

is the between-group variance of clinicians within the episode group

That is, reliability is calculated as the ratio of between-group variance to the sum of between-group variance and within-group variance. When estimating the within-group and between-group variation, conservative assumptions were used to account for the possible correlation between months for the same beneficiary when attributed to a single clinician or clinician group. 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

93.9 percent of TINs at the 10-beneficiary volume threshold, 100 percent of TINs at the 20 and 30 beneficiary volume threshold, and 100 percent of TIN-NPIs at 10, 20, and 30 beneficiary volume thresholds have mean reliability greater than or equal to 0.4. At a testing volume threshold of at least 10 beneficiaries, the mean reliability for TINs is 0.79 and for TIN-NPIs is 0.84. The mean reliability continues to increase at the 20 and 30-beneficiary volume thresholds.

Table 2: Reliability Results at Various Volume Thresholds

Volume Threshold (# of Beneficiaries)	TIN		TIN-NPI	
	Mean Reliability	% ≥ 0.4	Mean Reliability	% ≥ 0.4
10	0.79	93.9%	0.84	100.0%
20	0.82	100.0%	0.89	100.0%
30	0.84	100.0%	0.91	100.0%

3.2.4 Interpretation

Measure Reliability

Overall reliability of the TPCC measure is very high at a volume threshold of 20 beneficiaries or more for both TINs and TIN-NPIs due to the large number of beneficiary months attributed to clinicians. CMS generally considers 0.7 as the threshold indicating ‘high’ reliability, which is supported by previous work on reliability.⁹

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 TPCC measure was revised through a structured, iterative process for gathering detailed input from recognized clinician experts on the measure. These convened expert panels

⁹ 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

methodically assessed 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 attribution rules would effectively assign clinicians to patients whose primary care they are responsible for managing and over which they therefore would have reasonable influence (i.e., the ability of the measure score to differentiate good from poor performance).

In developing and refining this measure, Acumen incorporated input from a technical expert panel (TEP) and stakeholder feedback from national field testing. The TEP comprised 19 members from diverse backgrounds, including clinicians, healthcare providers, academia, and patient advocacy organizations. The TEP provided input on potential refinements to the measure through meetings in August 2017, May 2018, and November 2018. Based on input from the first two TEP meetings, TPCC attribution rules were refined to better identify care relationships and fairly attribute clinicians. For example, the timing of attribution was refined so that cost could no longer be assigned prior to seeing a beneficiary.

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 for the revised TPCC measure and feedback reports for attributed clinicians and clinician groups. During this period, 567,239 field test reports for TINs and TIN-NPIs were available for download and review for the revised TPCC measure. Following field testing and the November 2018 TEP meeting, a number of refinements were made to the measure including the addition of specialty exclusions to remove from attribution those clinicians belonging to specialties that are unlikely to be responsible for primary care.

Empirical Validity Testing

We undertook two approaches to estimate the measure's validity. In the first approach, we evaluated the empirical validity of the TPCC measure by examining differences in risk-adjusted cost for known indicators of resource or service utilization based on literature, specifically complications related to acute admission and post-acute care utilization. For this analysis, we compared the mean risk- and specialty-adjusted monthly cost for beneficiaries with and without complications related to acute admission and post-acute care utilization occurring in the measurement period. This analysis sought to confirm the expectation that the TPCC measure captures variation in service utilization.

In the second approach, we evaluated how different types of services impact risk-adjusted costs. We classified certain services included in the TPCC measure into clinically coherent groups of services, called service categories. The service categories are:

- Acute Inpatient Services
- Post-Acute Care - Home Health (HH)
- Post-Acute Care - Skilled Nursing Facility (SNF)
- Post-Acute Care - Inpatient Rehabilitation Facility (IRF) and Long-term Care Hospital (LTCH)
- Emergency Services Not Included In Hospital Admission (Non-Hospital Admission Emergency Services)
- Outpatient Evaluation And Management Services, Procedures, And Therapy

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. We calculated the Pearson correlation between the cost of each service category and the risk-adjusted cost.

We expected that at least some of the Post-Acute Care categories and the Acute Inpatient Services category would have the highest correlation with risk- and specialty-adjusted beneficiary month cost even after accounting for beneficiary characteristics, as these types of care are often associated with costly services related to treatment of complications.¹⁰

3.3.3 Statistical Results from Validity Testing

Table 3 shows results from the first analysis of validity for the distribution of risk- and specialty-adjusted monthly cost across beneficiary months for a beneficiary during the measurement period. The mean average risk- and specialty-adjusted monthly cost for a beneficiary during the measurement period is \$1,130. The mean average risk- and specialty-adjusted monthly cost *with* services relating to Acute Admissions is \$2,523, compared with \$812 for a beneficiary *without* services relating to this complication. The mean average risk- and specialty-adjusted monthly cost *with* services relating to Post-Acute Care is \$2,324, compared with \$940 for a beneficiary *without* services relating to complications.

Table 3: Distribution of Average Risk- and Specialty-Adjusted Monthly Cost

Cost Driver Category	Risk- and Specialty-Adjusted Cost										
	Mean	Std. Dev.	P1	P5	P10	P25	P50	P75	P90	P95	P99
Average Monthly Cost for Final Beneficiaries	\$1,130	\$1,474	\$0	\$87	\$142	\$290	\$640	\$1,441	\$2,634	\$3,634	\$6,843
Average Monthly Cost for Beneficiaries without Acute Admissions	\$812	\$1,080	\$0	\$74	\$123	\$243	\$488	\$973	\$1,813	\$2,567	\$5,020
Average Monthly Cost for Beneficiaries with Acute Admissions	\$2,523	\$2,048	\$410	\$660	\$853	\$1,318	\$2,040	\$3,046	\$4,528	\$5,923	\$11,004
Average Monthly Cost for Beneficiaries without Post-Acute Care (IRF, LTCH, HH, SN)	\$940	\$1,292	\$0	\$78	\$129	\$257	\$536	\$1,136	\$2,158	\$3,019	\$5,948
Average Monthly Cost for Beneficiaries with Post-Acute Care (IRF, LTCH, HH, SN)	\$2,324	\$1,923	\$206	\$440	\$632	\$1,104	\$1,897	\$2,934	\$4,353	\$5,610	\$9,791

The service categories analysis demonstrates the correlation between service categories and risk- adjusted cost at the TIN and TIN-NPI levels. At both the TIN and TIN-NPI levels, there is a strong correlation between the SNF service category and risk-adjusted cost (correlation: 0.53). At both the TIN and TIN-NPI levels there is a strong correlation between Outpatient E&M Services, Procedures, and Therapy and risk-adjusted cost (correlation: 0.44). At both the TIN

¹⁰ Khan, N.A., Quan, H., Bugar, J.M. et al., "Association of postoperative complications with hospital costs and length of stay in a tertiary care center" J Gen Intern Med (2006) 21: 177.

and TIN-NPI levels, there is a moderate correlation between the Acute Inpatient Services category and risk-adjusted cost (correlation: 0.38). In contrast, at the TIN and TIN-NPI levels the HH category has low correlation with risk-adjusted cost (correlation: 0.14). Similarly, the low correlation between the Non-Hospital Admission Emergency Services category and risk-adjusted cost (correlation: 0.16).

3.3.4 Interpretation

As expected, the average risk- and specialty-adjusted monthly costs for beneficiaries with acute inpatient admissions and post-acute care in the measurement period are higher than for beneficiaries without those services. This indicates that the measure may penalize clinicians who have higher rates of complications related to these types of services, while not disincentivizing the provision of appropriate care in other areas.

The results of the service category analysis demonstrate that the TPCC measure is able to accurately capture higher resource use across various types of services. Importantly, we see that the correlation with risk-adjusted cost is strong not only for high-cost categories such as Acute Inpatient Services (average cost for a clinician: \$9,373), but also for lower cost categories such as post-acute care for SNF services (average cost for a clinician: \$3,274). This indicates that the correlation does not come from a mechanical increase in beneficiary month costs from high-cost categories.

3.4 Exclusions Analysis

3.4.1 Method of Testing Exclusions

TPCC aims to measure the broader population of Medicare patients. Exclusions are primarily used to ensure that, as part of data processing, sufficient data are available to accurately determine resource use and calculate risk adjustment for each beneficiary. These exclusions, along with their rationales, are listed below:

- *The beneficiary was not continuously enrolled in Medicare Parts A and B unless partial enrollment was the result of either new enrollment or death only.*
 - These beneficiaries may have gaps in their Medicare claim records when benefits are covered by other payers.
- *The beneficiary resides outside the United States or its territories during the measurement period*
 - Differences in access to care and the reimbursement policy for healthcare services provided outside the U.S. can lead to unfair comparisons of cost.
- *The beneficiary receives benefits from the Railroad Retirement Board (RRB)*
 - Beneficiaries covered by the RRB may have healthcare benefits normally covered by Medicare paid by the RRB, which may bias the observed cost for these beneficiaries.

Given the rationales for these exclusions, we would expect these excluded beneficiaries' Medicare costs to have different rates and measurability than the included beneficiaries. For the exclusions, we examined annual Medicare Parts A and B spending from potentially attributable triggering events (i.e., candidate events) for excluded beneficiaries compared to spending for beneficiaries included in measure calculation to assess the differences between the two patient cohorts.

3.4.2 Statistical Results from Testing Exclusions

Table 4 below presents observed cost statistics from candidate events for beneficiaries who are excluded from the TPCC measure and from the set of final candidate events for beneficiaries included in the TPCC measure.

Table 4: Observed Cost for Measure Exclusions

Exclusion	Beneficiaries		Observed Cost		
	#	%	Mean	Percentile	
				10 th	90 th
No Continuous Enrollment in Medicare Parts A and B or Any Enrollment in Part C	5,399,782	15.15%	\$18,536	\$476	\$51,168
Beneficiary Resides Outside of U.S. or Territories	15,143	0.04%	\$11,903	\$274	\$34,403
Beneficiary Enrollment in Medicare for Railroad Workers and their Families	334,203	0.94%	\$18,500	\$1,168	\$50,283
Final Candidate Events after Beneficiary-level Exclusions	29,955,364	84.04%	\$19,276	\$1,320	\$51,805

3.4.3 Interpretation

The excluded beneficiary populations show lower annual Medicare cost.

The mean observed cost of candidate events for beneficiaries without continuous enrollment in Medicare Parts A and B is slightly lower compared to the mean observed cost for final candidate events. This difference is particularly pronounced at the 10th percentile. These results indicate that including these beneficiaries could bias the observed cost for these beneficiaries' candidate events. In addition to lowering potential bias, implementing exclusions also mitigates the extent to which gaps in a beneficiary's claims history can adversely affect the determination of risk factors for risk adjustment. For example, certain conditions that arise and should be reflected in the risk adjustment model may not be observed in the available claims data due to gaps in the beneficiary's Part A and B claims history.

Cases where the beneficiary resides outside the United States or its territories during the measurement period are associated with even lower costs. This difference from the final candidate events is pronounced at both tails of the distribution. This could be due to restricted access to care while outside of the U.S. and its territories. Exclusion of these beneficiaries is therefore justified for similar reasons as those provided above for beneficiaries who are not continuously enrolled in Medicare Parts A and B.

Beneficiaries that receive benefits from the RRB also have slightly lower mean cost compared to final candidate events after the exclusions are applied. While costs for candidate events for these beneficiaries do not substantially differ from the final set of candidate events, their Medicare claims history might have gaps in cases where their services are paid for by the RRB.

3.5 Risk Adjustment or Stratification

3.5.1 Method of Controlling for Differences

Differences in patient case mix are controlled for using separate CMS Hierarchical Condition Category Version 22 (CMS-HCC V22) models for new enrollees, continuing enrollees, enrollees in long-term institutional settings. In addition, for beneficiaries with ESRD, the CMS ESRD Version 21 (CMS-ESRD V21) models are used for new enrollees with ESRD, and community enrollees with ESRD. The CMS models were developed for use in the Medicare Advantage

program and the accuracy of the continual upkeep and performance of these models is reported to Congress every three years under the 21st century Cures Act.¹¹

The TPCC measure follows the CMS-HCC V22 risk adjustment models for new enrollee, community, and long-term institutional beneficiaries *without ESRD*. A beneficiary month is measured under the new enrollee model if they do not have a full one-year lookback of Medicare claims data as of the start of a beneficiary month. As a result, the model is derived primarily from beneficiary enrollment data. This model adjusts for gender, age, dual Medicare and Medicaid enrollment, and whether the beneficiary was originally entitled to Medicare due to disability through a series of interacted covariates. Beneficiaries with sufficient Medicare claims history are measured under the community or the institutional model if they are institutionalized in a long term care facility. In both models, severity of illness is measured using HCCs and disease interactions. 79 HCCs are accounted for under CMS-HCC V22 model for beneficiaries classified as community enrollees and long-term institutional enrollees while the exact number and types of disease interaction can vary. Both models interact beneficiary age with gender. In addition, the community model interacts dual enrollment status, gender, and the indicator for whether the beneficiary was originally entitled to Medicare due to disability, while the institutional model adjusts for disability as the original reason for Medicare enrollment and dual enrollment status independently.

For ESRD beneficiaries receiving dialysis, the TPCC measure utilizes the CMS-ESRD V21 risk adjustment models. Differentiated models are implemented for dialysis new enrollees and dialysis community enrollees. Similar to the CMS-HCC V22, enrollees are classified as new enrollees if they were not continuously enrolled in Parts A and B for the one-year lookback period prior to each beneficiary month. As a result of this, the model primarily uses information from the beneficiary's enrollment data. This model adjusts for gender, age, dual enrollment status, and whether the beneficiary was originally entitled to Medicare due to disability through a series of interacted covariates. In addition to accounting for these patient characteristics, the dialysis community model also risk adjusts for medical severity using 87 HCCs and additional disease interactions.

The CMS-ESRD V21 and CMS-HCC V22 models both generate a risk score for each beneficiary that summarizes the beneficiary's expected cost of care relative to other beneficiaries. Risk scores for ESRD beneficiaries are normalized to enable comparison with the HCC V22 risk scores. This is achieved by multiplying ESRD risk scores by the mean annual Medicare spending for the ESRD population applied in the CMS-ESRD V21 model and dividing by the mean annual Medicare spending for the total Medicare population applied in the CMS-HCC V22 model, effectively renormalizing ESRD risk score values to the equivalent scale of the HCC models. A risk score equal to one indicates risk associated with expenditures for the average beneficiary nationwide. Risk scores below or above one indicate below and above average risk, respectively.

Following the normalization of risk scores, observed costs for each beneficiary month are divided by the normalized risk score to obtain risk-adjusted monthly costs. These costs are then winsorized at the 99th percentile by assigning the 99th percentile of monthly costs to all attributed beneficiary months with costs above the 99th percentile. Finally, monthly costs are normalized to account for differences in expected costs based on the number of clinician groups to which a beneficiary is attributed in a given month. This normalization is applied by dividing

¹¹ CMS, "Report to Congress: Risk Adjustment in Medicare Advantage December 2018," <https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/RTC-Dec2018.pdf>

monthly costs by the cube root of the number of TINs to which a beneficiary is attributed for a particular a month.

Full details of the risk adjustment models and their respective risk adjustors are in the Measure Codes List.¹²

3.5.2 Conceptual, Clinical, and Statistical Methods

We selected the CMS-HCC V22 and CMS-ESRD V21 models based on previous studies evaluating their appropriateness for use in risk adjusting Medicare claims data. These models were developed specifically for use in the Medicare population, meaning that they account for conditions found in the Medicare population and are calibrated on Medicare fee-for-service beneficiaries. In addition, the CMS-HCC and ESRD models are routinely updated for changes in coding practices (e.g., the transition from ICD-9 to ICD-10 codes) and they are exhaustive on these code sets. Because the CMS-HCC and ESRD models have already been extensively tested, we reference this testing from the December 2018 CMS Report to Congress on Risk Adjustment in Medicare Advantage and focus any additional testing on how the CMS-HCC and ESRD models influence the final TPCC measure score.¹³

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). Testing results for factors included in the CMS-ESRD V21 2016 and CMS-HCC V22 2016 models can be found in the December 2018 CMS Report to Congress.¹⁴

3.5.5 Analyses and Interpretation in Selection of Social Risk Factors

The CMS-HCC V22 and CMS-ESRD V21 include risk adjustors for gender and low income status (as identified through dual enrollment status) in a series of interacted variables to account for social risk factors. The TPCC measure does not replicate these risk adjustment models, but instead uses the risk score coefficients obtained directly from CMS as calculated for use in the Medicare Advantage program. To test the impact of these social risk factors on the TPCC measure specially, a similar regression was conducted that analyzed mean monthly cost for beneficiaries included in the TPCC measure. Acumen examined the impact of including social risk factors in the risk adjustment model by running goodness of fit tests when comparing the replicated CMS-HCC V22 and CMS-ESRD V21 models with and without the inclusion of gender and dual enrollment status. Beneficiary gender and dual status were obtained from the EDB and CME.

First, we analyzed the model coefficients and p-values for each of the models with and without social risk factors to understand whether any of the social risk factor covariates are predictive of

¹²CMS, "Total Per Capita Cost 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-TPCC-measure-specs.zip>.

¹³ CMS, "Report to Congress: Risk Adjustment in Medicare Advantage, December 2018," <https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/RTC-Dec2018.pdf>

¹⁴ Ibid.

a beneficiary's monthly cost. The T-test revealed many significant p-values, indicating that social risk factors are likely predictive of resource use among beneficiaries for the relevant characteristic. The adjusted R-squared of the replicated models without social risk factors decreased by less than 0.002 for the Community, Institutional, New Enrollee, and Community Dialysis models. The adjusted R-squared for the New Enrollee Dialysis model decreased by 0.027.

Secondly, we analyzed the impact of social risk factors on the overall model performance by looking at differences in measure scores calculated with and without social risk factors. Results indicate minor differences in measure score performance, even for providers at high or low extremes of risk. The measure scores for 89.6 percent of TINs and 87.9 percent of TIN-NPIs changed by ± 1 percent or less when social risk factors were added to the risk adjustment model. Scores for nearly 100 percent of TINs and TIN-NPIs changed by ± 10 percent or less. These results suggest that the effects of social risk factors on the model performance are likely captured through existing risk adjustment variables.

In addition, we analyzed the correlation between measure scores calculated with and without indicators for sex and dual enrollment status. The measure scores were highly correlated at both TIN and TIN-NPI levels, with Spearman correlation coefficients of 0.998 for both levels. These results indicate that the inclusion of social risk factors in the current risk adjustment model has a minor effect on measure scores.

3.5.6 Method for Statistical Model or Stratification Development

The R-squared values reported in the December 2018 CMS Report to Congress confirm the validity of the CMS-HCC V22 model for community enrollees.¹⁵

3.5.7 Statistical Risk Model Discrimination Statistics

Discrimination testing for the CMS-HCC model can be found in the studies by Pope et al. 2011 and more recent analyses shared in the December 2018 CMS Report to Congress.^{16,17}

3.5.8 Statistical Risk Model Calibration Statistics

Included in the standard testing of the HCC and ESRD models in the December 2018 CMS Report to Congress is calibration analyses interpreted as how accurately the risk models' predictions match the actual beneficiary cost. For each of the risk factors included in the models, predictive ratios were calculated as a ratio of predicted cost to actual cost for sub-groups of beneficiaries within the model sample to demonstrate the models' prediction accuracy. For all models, the predictive ratio is equal to or close to one across all risk factors, indicating that the model is accurately predicting actual beneficiary cost for that risk factor.

3.5.9 Statistical Risk Model Calibration – Risk Decile

Results of analyses examining predictive ratios by risk decile are included in the December 2018 CMS Report to Congress. Analyses of predictive ratio by risk decile assess the stability of the risk adjustment model among beneficiaries of similar case mixes. As shown in the December 2018 Report to Congress, analyses of these risk deciles for the measure shows that the predictive ratios are generally close to one across all risk score deciles.

¹⁵ CMS, "Report to Congress: Risk Adjustment in Medicare Advantage, December 2018," <https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/RTC-Dec2018.pdf>

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.

¹⁷ Ibid, 15.

3.5.10 Interpretation

As reported in Sections 3.5.8 and 3.5.9, the predictive ratios for each risk factor included in the model and 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 observed spending, regardless of individual risk factors or overall risk level.

3.6 Identification of Meaningful Differences in Performance

3.6.1 Method

Our method of determining clinically meaningful differences in measure scores consists of stratifying the clinician measure scores by defining 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 beneficiary months 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 meaningfully determine a difference in performance. In addition, this 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 TPCC measure:

- (i) the 99th percentile of the measure score is between 3.5 to 4 times the 1st percentile at both the TIN level and TIN-NPI levels;
- (ii) the TPCC measure score at the 90th percentile is approximately 70 percent greater than the score at the 10th percentile at both the TIN and TIN-NPI levels.

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 mean scores for clinicians across nine census divisions (excluding 'Unknown') are within a less than \$130 range (i.e., \$1,007 to \$1,137 at the TIN level and \$1,081 to \$1,173 at the TIN-NPI level). Similarly, clinicians in urban areas seem to perform comparably to those in rural areas on average (i.e., \$1,054 in urban compared to \$1,070 at the TIN level and \$1,116 in urban compared to \$1,108 in rural at the TIN-NPI level).

In terms of other clinician characteristics, analysis of clinicians by number of beneficiary months indicates that clinicians with more beneficiaries perform similarly to those responsible for fewer beneficiaries with a difference in mean score less than \$88 at both the TIN and TIN-NPI levels. 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 \$1,026 to \$1,141 and a range in mean TIN-NPI score of \$1,074 to \$1,233, indicating that the risk adjustment model is overall functioning as intended. Full results can be seen in the updated National Summary Data Report (NSDR).¹⁸

¹⁸ 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*,

3.6.3 Interpretation

There is clinically and practically significant variation in TPCC 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. Overall, as expected, results show that clinicians are not being systematically penalized or rewarded based on their patient case mix given the current TPCC measure design (i.e., the differences in cost measure scores are not because of the risk profile of the patient cohort).

3.7 Missing Data Analysis and Minimizing Bias

3.7.1 Method

Since CMS uses Medicare claims data to calculate the TPCC measure, Acumen expects a high degree of data completeness. To ensure further that we have complete and accurate data Acumen excludes beneficiaries enrolled in Medicare Part C or who have a primary payer other than Medicare during the measurement period. In such situations, Medicare Parts A and B claims data may not contain sufficient information to capture the beneficiary's complete clinical risk profile, which is required for 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 three categories of missing data which caused beneficiaries to be excluded from the TPCC measure. Frequency is presented in terms of the number of beneficiaries excluded due to missing data, as well as the number of TINs and TIN-NPIs who had at least one beneficiary excluded due to missing data. The missing data exclusions are:

- Beneficiary was not enrolled in Medicare Parts A and B, or was enrolled in Part C, during the measurement period
- Beneficiary resides outside of the U.S. or Territories
- Beneficiary Enrollment in Medicare for Railroad Workers and their Families

Table 5: Missing Data Categories for the TPCC Measure

Exclusion	# Beneficiaries	# TINs	# TIN-NPIs
No Continuous Enrollment in Medicare Parts A and B, and Any Enrollment in Part C	5,399,782	161,262	781,814
Beneficiary Resides Outside of U.S. or Territories	15,143	18,172	39,162
Beneficiary Enrollment in Medicare for Railroad Workers and their Families	334,203	84,408	383,372

3.7.3 Interpretation

As the TPCC measure is calculated with Medicare claims data, Acumen expects a high degree of data completeness for those beneficiaries with Medicare Part A and B coverage and removes

<https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/value-based-programs/macra-mips-and-apms/macra-feedback.html>.

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 the measurement period.

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 conducted in accordance with that program’s reporting practices.

4.3.1.2 Missing Data

This measure requires complete beneficiary information, and a small number of beneficiaries with missing data are excluded to ensure completeness of data and accurate comparability across beneficiary months.

4.3.1.3 Sampling

n/a.

5.0 Usability and Use

5.1 Use

5.1.1 Current and Planned Use

The revised measure was re-evaluated for potential use in the 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 revised TPCC and MSPB clinician measures, developed during 2018 for a 35-day comment period (October 3 to November 5, 2018). We provided TPCC field test reports to a sample of eligible clinician groups and clinicians. Each report included information on measure performance for a clinician or clinician group attributed 20 or more beneficiaries.¹⁹ The testing sample was selected to balance coverage and reliability, since a key goal of field testing was to test the measure 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.

- TPCC: 567,239; 120,266 TINs; 446,973 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 for the revised TPCC measure 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.²⁰ 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, 12,902 field test reports for TPCC were downloaded by 703 clinician groups (TINs) and 12,199 clinicians (TIN-NPIs). Stakeholder comments from field testing were summarized for the TEP to consider in recommending refinements to the measure based on the testing data and feedback.

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

Data Provided During Field Testing

Each TPCC field test report contained the following:

- The clinician or clinician group TPCC Field Test Report Measure Score along with the national median score and percentile rank

¹⁹ The field test reports were available for download from the CMS Enterprise Portal: <https://portal.cms.gov/wps/portal/unauthportal/home/>.

²⁰ 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>.

- TPCC cost breakdown by claim type to explain the factors driving the clinician or clinician group measure score (e.g., home health agency, hospice, inpatient, outpatient)
- TPCC cost breakdown by specialty type. The TPCC measure is mostly attributed to primary care physicians and non-physician practitioners, so figures for these two categories are further broken down by specialty (e.g., general practice, family practice, internal medicine, geriatric medicine)
- TPCC cost breakdown by categories of service to show the average cost per category (e.g., acute inpatient services, post-acute care)
- Statistics of the TIN or TIN-NPI's specific performance compared to the state and national average (e.g., number of beneficiaries, average standardized cost per beneficiary)

A mock field test report can be viewed on the CMS MACRA Feedback webpage.²¹ Along with the Field Test Report, attributed clinicians and clinician groups received a beneficiary-level CSV file that include the risk profile of the beneficiaries attributed.

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.

Acumen worked with the Physician Value helpdesk and QPP Service Center to answer stakeholder questions during field testing and continued to answer questions after the feedback period ended.

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.²²

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.

²¹ CMS, "Total Per Capita Cost 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-TPCC.pdf>.

²² CMS, Webinar Recordings, Slides and Transcripts, *QPP Webinar Library* <https://qpp.cms.gov/about/webinars>.

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 12 comments on the revised TPCC 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 the preliminary recommendation, which received seven comments specific to the TPCC measure.²³ These public comment periods were facilitated by NQF. Stakeholders were able to submit their comments via the NQF website.

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:

- *Stakeholder engagement and involvement remains an important aspect of the measure development process.* Stakeholders expressed appreciation for the opportunity to provide feedback during field testing and for CMS' continued efforts to involve them in the measure development process. Commenters also valued the decision to operationalize previously collected feedback, as demonstrated through the addition of measure-specific workgroups to the development process.
- *Field test reports present useful information for understanding clinician performance, though reduced complexity could encourage more clinician participation.* Stakeholders praised the presentation and content of the field test reports. However, the complexity of the information presented in the reports was a challenge for some stakeholders.
- *Improved supplemental field testing materials are helpful but can be further refined.* Some stakeholders found the supplemental field testing materials to be informative and thorough, providing useful information on field testing and the specifications of the cost measures. However, many noted that although the materials are comprehensive, they remain lengthy and complex, and they believe the amount of information provided is too overwhelming to be useful.
- *Ample time for review of field testing reports and materials is vital to collecting meaningful stakeholder feedback.* Some stakeholders suggested the field testing period be extended or kept open, given the large amount and complexity of the information that was presented.
- *Field test report access continues to present challenges for stakeholders.* Some stakeholders noted that they faced difficulties creating accounts and downloading their field test reports from the CMS Enterprise Portal and these challenges may have negatively impacted the number of clinicians that were able to participate in field testing. Stakeholders urged CMS to communicate directly with clinicians receiving field test reports and to find an alternative for delivering and accessing the reports.

The report additionally contains measure-specific feedback, which was used as the basis for the post-field testing refinements that were made to the measures, summarized below:

- Refinements to the list of primary care services used as candidate events to ensure they better reflect primary care services

²³ Measure Applications Partnership, *National Quality Forum*.

https://www.qualityforum.org/Setting_Priorities/Partnership/Measure_Applications_Partnership.aspx.

- Addition of the specialty exclusions so that HCFA specialties who are not identified to be reasonably responsible for providing primary care are not attributed the TPCC measure
- Ensuring a cost adjustment is applied to account for costs that vary across specialties and across TINs with varying specialty compositions

5.1.2.5 Feedback from Other Users

Pre-Rulemaking

The revised TPCC measure underwent MAP review during the 2018-2019 cycle. In December 2018, the MAP Clinician Workgroup gave the preliminary recommendation of 'conditional support for rulemaking,' with the condition of NQF endorsement. In January 2019, the MAP Coordinating Committee reversed the Clinician Workgroup's preliminary recommendation and provided a final recommendation of 'do not support for rulemaking with potential for mitigation'. More detail on the mitigating factors is available in the MAP's final report.²⁴ While the measure did not receive MAP support due to their concerns regarding the revised specifications, CMS believes that the revised measure provides a more appropriate and valid attribution approach than the current TPCC measure used in MIPS and has adequately addressed the mitigating factors outlined by the MAP. For example, CMS has engaged in a range of education and outreach to increase familiarity with the revisions to the measure, including through field testing and national webinars both during and after field testing. The measure has also been tested, including examining how the measure performs at small numbers, and has been found reliable for TINs at various sizes. Testing results, including this MJF, are also publicly posted on the MACRA Feedback Page.²⁵

5.1.2.6 Consideration of Feedback

Field Testing

Careful consideration was given to all feedback gathered during field testing, and several updates were made to the measure based on the recommendations of field testing commenters and TEP comprised of subject matter and measure-development experts.

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 TEP, 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.

The changes to the TPCC measure made after consideration of field testing analyses and stakeholder feedback are:

- **Candidate events:** Primary care services list was refined to better reflect primary care services, and went from around 5200 codes to 3200 codes. The categories for primary care services have not changed.
- **Attributable Clinicians:** Included HCFA specialties to determine include or exclude clinicians from attribution:
 - HCFA specialties eligible for attribution are those that can be reasonably be responsible for providing primary care:

²⁴ "MAP Clinicians 2019 Considerations for Implementing Measures Final Report," *National Quality Forum*, http://www.qualityforum.org/Publications/2019/03/MAP_Clinicians_2019_Considerations_for_Implementing_Measures_Final_Report.aspx.

²⁵ 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>.

- Primary care specialties
- Internal medicine sub-specialties that frequently manage chronic patients with significant conditions in their areas of specialties along with other medical comorbidities
- Non-physician clinicians who often provide primary care services
- HCFA specialties excluded from attribution were identified as not providing chronic care for significant medical conditions and fall into the following broad categories:
 - Surgical sub-specialties
 - Non-physicians without chronic management of significant medical conditions
 - Internal medicine sub-specialties with additional highly procedural sub-specialization
 - Internal medicine that practice primarily inpatient without chronic management
 - Pediatricians who do not typically practice adult medicine
- **Specialty Adjustment:** Will be applied based on clinician specialty.

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 2017.²⁶ 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.

²⁶ CMS, "2017 Quality Payment Program Reporting Experience," <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 are currently no related NQF-endorsed or non-NQF-endorsed cost measures that address this same measure focus or target population. There are no competing NQF-endorsed or non-endorsed cost measures that address both this same measure focus *and* at this same target population.

6.2 Harmonization

n/a

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.