Quality Payment
PROGRAM

Inpatient (IP) Percutaneous Coronary Intervention (PCI)

Measure Testing Form

February 2023

Contents

1.0 Introduction

This Measure Testing Form (MTF) provides a brief summary of the preliminary measure testing results as part of the comprehensive reevaluation process for three episode-based cost measures. Readers may review these results, alongside other documentation, to provide feedback on the measure using the comprehensive reevaluation survey. The testing results reflect both the version of the measure that is currently in-use in MIPS and a revised version of the measure that is undergoing updates potential use in MIPS in future years. Please see the Draft Cost Measure Methodology for a description of the measure specifications and the Draft Measure Codes List for the list of codes used to specify the measure.[1]

1.1 Project Title and Overview

The Centers for Medicare & Medicaid Services (CMS) has contracted with Acumen, LLC to develop and maintain episode-based cost measures for potential use in the Merit-Based Incentive Payment System (MIPS) to meet the requirements of the Medicare Access and CHIP Reauthorization Act (MACRA) of 2015. The contract name is "Physician Cost Measures and Patient Relationship Codes (PCMP)." The contract number is 75FCMC18D0015, Task Order 75FCMC19F0004.

1.2 Measure Name

Inpatient (IP) Percutaneous Coronary Intervention (PCI) Episode-based Cost Measure

1.3 Type of Measure

Cost/Resource Use

1.4 Data

The study period is January 1, 2021 through December 31, 2021. All episodes ending during the study period that meet inclusion and exclusion criteria are included in testing. The measure is calculated with Medicare Parts A and B, administrative claims data, Long-Term Minimum Data Set, Medicare Enrollment Database. For testing purpose, other data sources are used, including the American Community Survey, Common Medicare Environment.

Testing results are presented at a testing volume threshold of 20 episodes for clinician groups and individual practitioners. Clinician groups are identified by a Tax Identification Number (TIN). Individual clinicians are identified using a combination of a Tax Identification Number and National Provider Identifier (TIN-NPI).

[1]These documents will be available on the MACRA Feedback Page once field testing begins. https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback

2.0 Preliminary Testing Results

This section presents preliminary testing results based on the revised measure as specified for the public comment period. Section 2.1 provides an overview of changes in the measure coverage, clinician population, and reliability between the current measure and the revised version of the measure. Sections 2.2 and 2.3 show additional evidence of scientific acceptability of the measure. Section 2.4 presents empirical results of the risk adjustment and stratification methods used by this measure. Section 2.5 examines the impact of adding social risk factors to the measure's risk adjustment model. Lastly, Section 2.6 examines the impact of exclusion criteria used by the measure through their frequency and resource use patterns.

2.1 Impacts of Revisions to the Measure

2.1.1 Measure Coverage

Table 1 shows the number of beneficiaries covered by this measure. Table 2 shows the characteristics of TINs and TIN-NPIs who are attributed at least 20 episodes. Compared to the current MIPS version, the revised version captures more beneficiaries.

Table 1: Measure Coverage

Metric	Value	
	Current MIPS Measure	Revised Measure
Number of Beneficiaries	18,640	73,872
Mean Age	73.7	74.4
Female %	37.6%	38.0%

Table 2: Clinician Characteristics

Metric	TIN				TIN-NPI			
	Current MIPS Measure		Revised Measure		Current MIPS Measure		Revised Measure	
	Count	%	Count	%	Count	%	Count	%
Count	276	100%	1,293	100%	-	-	472	100%
Number of Episodes Attributed	-	-	-	-	-	-	-	-
20-39 Episodes	201	72.8%	595	46.0%	-	-	445	94.3%
40-59 Episodes	59	21.4%	273	21.1%	-	-	24	5.1%
60-79 Episodes	9	3.3%	155	12.0%	-	-	2	0.4%
80-99 Episodes	5	1.8%	87	6.7%	-	-	0	0.0%
100-199 Episodes	2	0.7%	146	11.3%	-	-	1	0.2%

Metric	TIN				TIN-NPI			
	Current MIPS Measure		Revised Measure		Current MIPS Measure		Revised Measure	
	Count	%	Count	%	Count	%	Count	%
200-299 Episodes	0	0.0%	29	2.2%	-	-	0	0.0%
300+ Episodes	0	0.0%	8	0.6%	-	-	0	0.0%
Census Region	-	-	-	-	-	-	-	-
Northeast	48	17.4%	182	14.1%	-	-	66	14.0%
Midwest	67	24.3%	325	25.1%	-	-	130	27.5%
South	114	41.3%	543	42.0%	-	-	209	44.3%
West	47	17.0%	243	18.8%	-	-	67	14.2%
Unknown	0	0.0%	0	0.0%	-	-	0	0.0%

NOTE: The current MIPS version of this measure is only scored at the group-level (i.e., TIN-level) because there are no individual clinicians (i.e., TIN-NPIs) meeting the 20-case minimum.

2.1.2 Frequently Attributed Specialties

Table 3 shows the top 5 attributed specialties for the revised measure, using a 20-episode testing volume threshold. The most frequently attributed specialties reflect the intent of the measure to capture costs of percutaneous coronary intervention procedures, including cardiologists, internists, and related specialists.

Table 3: Count of the Top 5 Attributed Specialties

Specialty	Number of TIN-NPIs Attributed
Cardiology	210
Interventional Cardiology	150
Nurse Practitioner	41
Internal Medicine	31
Physician Assistant	21

2.1.3 Reliability

Reliability evaluates a measure's ability to consistently differentiate the performance of one clinician from another. The signal-to-noise ratio is used to estimate reliability, which indicates how much of the variation in the measure score is explained by differences among clinicians' performance (i.e., signal) instead of differences within each clinician's performance (i.e., noise). Specifically, noise is the variation from one episode to another during the performance period for a particular clinician.

Table 4 shows reliability metrics at various testing volume thresholds. While higher thresholds yield higher reliability results, it is at the cost of further reducing the number of clinicians and clinician groups eligible for the measure, which would reduce the potential impact of the

measure. For the purposes of testing, we used a 20-episode volume threshold (bolded in the table below) to align with the current MIPS version. If the measure is implemented in the MIPS in the future, CMS will establish a case minimum through notice-and-comment rulemaking.

Table 4: Sample Size, Mean Reliability, and Proportion of Clinicians above Moderate Reliability at Various Testing Volume Thresholds

Version	Testing Volume Threshold	TIN			TIN-NPI		
		Number of TINs	Mean Reliability	Percent Above 0.4	Number TIN-NPIs	Mean Reliability	Percent Above 0.4
Current Measure	10	731	0.39	37.8%	166	0.39	27.1%
Revised Measure	10	1,949	0.47	56.8%	3,590	0.35	18.8%
Current Measure	**20**	**276**	**0.52**	**100.0%**	**-**	**-**	**-**
Revised Measure	**20**	**1,293**	**0.56**	**85.5%**	**472**	**0.49**	**100.0%**
Current Measure	30	144	0.58	100.0%	-	-	-
Revised Measure	30	949	0.62	100.0%	85	0.59	100.0%

NOTE: The current MIPS version of this measure is only scored at the group-level (i.e., TIN-level) because there are no individual clinicians (i.e., TIN-NPIs) meeting the 20-case minimum.

At the testing volume of 20 episodes, the revised version shows moderate reliability, specifically 0.56 at the TIN level and 0.49 at the TIN-NPI level (Table 4). CMS generally considers 0.4 as the threshold indicating 'moderate' reliability and 0.7 indicating 'high' reliability, which is supported by previous work into reliability and the threshold was finalized in the 2022 Physician Fee Schedule final rule.[2,3] The majority of TINs and TIN-NPIs meet or exceed the moderate reliability threshold of 0.4 at the 20-episode testing volume threshold, 85.5% at the TIN level and 100.0% at the TIN-NPI level.

Compared to current measure, the number of clinician groups eligible for the revised measure is over four-fold the number eligible for the current measure, with an increase in reliability. Likewise, the current measure has no providers at the TIN-NPI reporter level meeting the 20-episode testing volume threshold. In contrast, the revised measure captures 472 TIN-NPIs with at least 20 episodes.

2.2 Validity

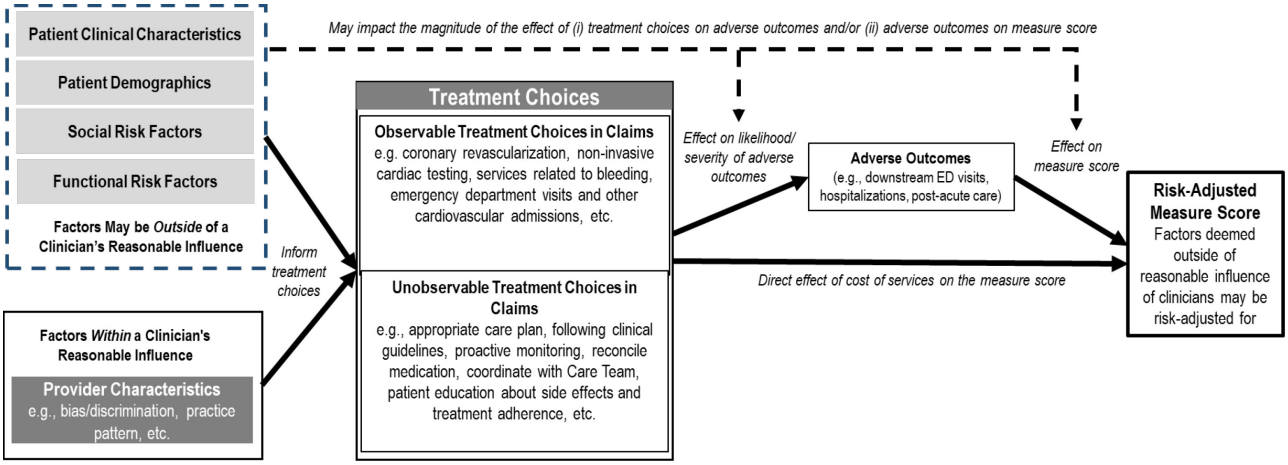
Validity is a criterion that evaluates whether the cost measure is able to quantify the construct that it aims to measure, which is the cost directly related to treatment choices and cost of adverse outcomes as a result of care. Validity is evaluated empirically by estimating the effect of

[2] 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

[3] CMS, "Medicare Program; CY 2022 Payment Policies Under the Physician Fee Schedule and Other Changes to Part B Payment Policies; Medicare Shared Savings Program Requirements; Provider Enrollment Regulation Updates; and Provider and Supplier Prepayment and Post-Payment Medical Review Requirements," 86 FR 64996-66031

relevant treatment choices on the measure score using multiple regression, based on the conceptual model outlined in Figure 1.

Figure 1: Conceptual Model of the Relationship between Treatment Choices and the Measure Score



The cost measure is designed to reflect the cost directly related to treatment choices, as well as the cost of adverse outcomes as a result of care. Therefore, treatment choices, either observable in claims or otherwise, by an attributed clinician can directly impact the measure score or indirectly when they're mediated through the cost of adverse outcomes. The cost of adverse outcomes, in turn, contributes to the total costs that are captured by the measure score.

To demonstrate that the measure score is reflective of both the direct and indirect effects of treatment choices, this analysis first estimates the association between treatment choices and the measure score while controlling for the cost of adverse outcomes. Then, the association between treatment choices and the cost of adverse outcomes is estimated to demonstrate the indirect effect.

Generally, adverse outcomes are non-trigger inpatient hospitalizations, non-trigger emergency room visits, and post-acute care. The remaining service categories are generally considered treatment. For each of these categories, the regression models use the mean cost across episodes that were attributed to an individual clinician. The measure score is represented by a clinician's mean observed cost over expected cost ratio across their attributed episodes.

Overall, the results demonstrate that the cost measure is reflective of both the cost directly related to treatment choices, as well as cost of adverse outcomes as a result of care (Table 5). Therefore, there's evidence that the measure is capturing what it purports to measure.

Model 1 shows that adverse events (e.g., readmissions, ED visits) and outpatient procedures observed after the trigger inpatient stay are associated with a worse measure score. Additionally, laboratory tests and durable medical equipment (DMEs) are associated with worse measure scores at the TIN-reporting level, and inpatient hospital services, physician services, and chemotherapy and other Part B-covered drugs are associated with worse measure scores at both reporting levels. Model 2 shows that outpatient therapy services (i.e., physical therapy, occupational therapy, and speech and language therapy) are associated with a lower cost of adverse events at the TIN reporting level.

Altogether, both models demonstrate that opportunities for improvement may include reducing the additional post-discharge procedures, readmissions, and ED visits. On the other hand,

ensuring patients receive adequate follow-up care, such as outpatient therapy services, is still important in reducing risk of adverse events.

Table 5: Estimated Effect of Treatment Choices (Revised Measure)

Categories of Service	Coefficient in Thousands [95% Confidence Interval] (p-value)			
	TIN		TIN-NPI	
	Model 1: Mean O/E = Mean Cost of Treatment Choices + Mean Cost of Adverse Events	Model 2: Mean Cost of Adverse Events = Mean Cost of Treatment Choices	Model 1: Mean O/E = Mean Cost of Treatment Choices + Mean Cost of Adverse Events	Model 2: Mean Cost of Adverse Events = Mean Cost of Treatment Choices
Adverse Events	0.04 [0.04,0.04] (p < 0.01)	-	0.03 [0.03,0.04] (p < 0.01)	-
Major Procedures	0.03 [0.01,0.05] (p = 0.01)	-0.20 [-0.74,0.34] (p = 0.47)	0.02 [-0.03,0.07] (p = 0.41)	-0.49 [-1.55,0.56] (p = 0.36)
Ambulatory/Minor Procedures	0.05 [0.05,0.06] (p < 0.01)	-0.07 [-0.24,0.11] (p = 0.43)	0.05 [0.03,0.06] (p < 0.01)	-0.12 [-0.43,0.20] (p = 0.47)
Outpatient Physical, Occupational, or Speech and Language Pathology Therapy	0.03 [0.00,0.06] (p = 0.09)	-0.98 [-1.71,-0.25] (p < 0.01)	-0.09 [-0.15,-0.03] (p < 0.01)	-0.46 [-1.82,0.90] (p = 0.50)
Laboratory, Pathology, and Other Tests	0.09 [0.03,0.14] (p < 0.01)	1.92 [0.67,3.17] (p < 0.01)	0.05 [-0.05,0.14] (p = 0.32)	3.75 [1.62,5.89] (p < 0.01)
Durable Medical Equipment and Supplies	0.03 [0.01,0.05] (p < 0.01)	0.40 [0.01,0.79] (p = 0.05)	0.02 [-0.01,0.04] (p = 0.27)	0.47 [-0.14,1.08] (p = 0.13)
Inpatient Hospital Trigger	0.01 [0.01,0.01] (p < 0.01)	0.06 [0.02,0.11] (p < 0.01)	0.01 [0.01,0.02] (p < 0.01)	0.08 [0.01,0.16] (p = 0.03)
Physician Services During Hospitalization Trigger	0.02 [0.01,0.02] (p < 0.01)	0.31 [0.14,0.48] (p < 0.01)	0.02 [0.00,0.03] (p = 0.01)	0.35 [0.07,0.64] (p = 0.02)
Chemotherapy and Other Part B-Covered Drugs	0.04 [0.03,0.05] (p < 0.01)	1.49 [1.22,1.77] (p < 0.01)	0.04 [0.03,0.06] (p < 0.01)	1.26 [0.87,1.65] (p < 0.01)

2.3 Performance Gap

Table 6 shows the distribution of the revised measure scores for clinicians and clinician groups. These results align with expectations based on our review of the literature and demonstrate that there is a performance gap in cost measure performance at both the clinician and clinician

group levels. The variation in the measure score, indicated by the interquartile range and standard deviation, is in the thousands of dollars. The results suggest that there is opportunity for improvement in performance across providers.

Table 6: Distribution of the Measure Score (Revised Measure)

Metric	TIN	TIN-NPI
Mean Score	$19,850	$21,762
Score Interquartile Range (IQR)	$1,245	$1,802
Standard Deviation	$1,020	$1,372
Coefficient of Variation	0.05	0.06
Score Percentile		
10th	$18,662	$20,134
25th	$19,189	$20,798
50th	$19,751	$21,608
75th	$20,434	$22,600
90th	$21,160	$23,644

2.4 Risk Adjustment and Stratification

Figure 1 shows the conceptual model that outlines how patient-level and clinician-level factors can influence the measure score, which is informed by both published external research and our own data analysis.[4,5,6,7,8] The conceptual model includes risk factors that are either known by the literature or informed by the initial and reconvened Clinical Expert Workgroups to be within or outside of the influence of the attributed clinician. Risk factors, including social risk factors (SRFs), can both influence the treatment choices and impact the size of the effect of treatment choices by mitigating the risk of adverse outcomes and the cost of adverse outcomes.

A systematic approach then guides the decision of which factors to include in the risk adjustment model. First, during initial development of the current MIPS measure, we reviewed the literature to gather known risk factors and drivers of resource use. These factors are usually diagnoses; therefore, the first set of risk adjustors are commonly the Hierarchical Condition Categories. Then, we consulted our clinical expert panels on additional factors that are known to be associated with resource use. Together with our clinical expert panel, we reviewed the stratified results on episode cost across many different patient characteristics. We arrived at the final list of risk adjustors used in the current MIPS measure based on those discussions and consensus among the clinical experts. We also reviewed literature and gathered additional input

[4]Centers for Medicare & Medicaid (CMS), Office of Minority Health. "Utilization of Z Codes for Social Determinants of Health among Medicare Fee-for-Service Beneficiaries." (2019) https://www.cms.gov/files/document/z-codes-data-highlight.pdf
[5]Assistant Secretary of Health and Human Services for Planning and Evaluation. Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs. Washington, D.C. December 2016.
[6]Chen LM, Epstein AM, Orav EJ, Filice CE, Samson LW, Joynt Maddox KE. Association of Practice-Level Social and Medical Risk With Performance in the Medicare Physician Value-Based Payment Modifier Program. JAMA. 2017;318(5):453-461.
[7]Medicare Payment Advisory Commission. Beneficiaries Dually Eligible for Medicare and Medicaid. 2018; https://www.macpac.gov/publication/data-book-beneficiaries-dually-eligible-for-medicare-and-medicaid-3/.
[8]Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health & Human Services. Second Report to Congress on Social Risk Factors and Performance in Medicare's Value-Based Purchasing Program. 2020. https://aspe.hhs.gov/reports/second-report-congress-social-risk-medicares-value-based-purchasing-programs

from the reconvened Clinician Expert Workgroup to determine whether revisions should be made to the risk adjustment model. Additionally, during our testing phases, we also follow a structured and systematic approach to decide whether SRFs should be risk-adjusted for, which is further described in Section 2.5.

2.4.1 Discrimination

Discrimination is a statistical criterion that evaluates the measure's ability to distinguish high-cost episodes from low-cost episodes, or the ability to explain the variance in cost of individual episodes. The amount of variance explained is estimated by the R-squared metric with the range between 0 and 1. The R-square value for the measure is 0.363, and 0.360 after adjusting for the model's complexity based on the number of risk adjustors used. In other words, 36.0% of the variation in the actual observed cost of episodes is explained by the risk adjustment model and sub-group stratification.

The remaining unexplained variance is due to variation in factors that are not adjusted for by the measure, such as the clinician's performance. The objective of a cost measure is to evaluate and differentiate the performance of clinicians. Therefore, achieving high explained variance is not essential because not all of the variation in cost of care should be adjusted. In collaboration with the experts from our clinical workgroup, this measure only adjusts for factors that are deemed to be outside of the influence of clinicians. Please see the Draft Cost Measure Methodology for more information on the full list of risk adjustors and sub-groups.

2.4.2 Calibration

Calibration evaluates the consistency of the measure in estimating episode cost across the full range of resource use patterns in the population. Calibration is estimated by the average predictive ratios across groups within the population, specifically groups are partitioned by deciles of expected episode cost. The predictive ratio is calculated using the formula of average expected cost / average observed cost for all episodes in each decile. A well-calibrated measure should have predictive ratios close to 1.00 across all deciles. In other words, such results show that the measure is consistent because it does not under- or over-predict cost throughout the range of resource use patterns in the population.

Table 7 shows that the model has consistent predictive ratios across risk score deciles, with each decile having a predictive ratio between 0.98 and 1.02. The average predictive ratio for all risk deciles is 1.00, which demonstrates that the risk adjustment does not under- or over- predict across the full range of resource use patterns in the population.

Table 7: Predictive Ratio by Decile of Predicted Episode Cost (Revised Measure)

Decile	Average Predictive Ratio
Decile 1	0.98
Decile 2	1.00
Decile 3	1.00
Decile 4	1.01
Decile 5	1.00
Decile 6	1.01
Decile 7	1.02
Decile 8	1.01
Decile 9	1.00

Decile	Average Predictive Ratio
Decile 10	0.99

2.5 Social Risk Factor Analysis

Beyond clinical characteristics of patients, the cost of care may be influenced by non-clinical factors related to a patient's social risk factors (SRFs), such as race, income, education, and employment. At the program level, MIPS adjusts for SRFs using the MIPS Complex Patient Bonus to ensure clinicians or groups treating more complex patients are not disadvantaged.[9] At the measure-level, the testing helps to navigate the tension between ensuring fairness for clinicians treating higher shares of vulnerable patients and the possibility of masking poor performance and perpetuating disparity if clinicians are held to different standards.

Table 8 outlines variables that may indicate SRFs and their advantages and disadvantages as indicators of individual-level SRFs. Based on availability of data, this analysis tested all variables except for the ICD-10 Z codes.

Table 8: Social Risk Factors Available for Analysis (Revised Measure)

Variable	Advantages	Disadvantages	Used in Testing
Dual Medicare and Medicaid enrollment status	• Available for all beneficiaries • Most powerful predictor of poor outcomes[10]	• Variation in Medicaid eligibility across states	Yes
Race/Ethnicity	• Available for most beneficiaries, except for ambiguous categories of "Unknown" or "Other"	• Social risk driven by someone's race is often correlated with and partially captured by dual status[11] • Only 5 categories available, which may lack granularity to fully capture disparities[12,13]	Yes
ICD-10 Z codes for social determinants of health	• Reflects individual-level factors that influence health status and contact with health services	• Not routinely and consistently coded on claims, only available for 0.1% of all fee-for-service claims in 2019[14]	No

[9] https://qpp-cm-prod-content.s3.amazonaws.com/uploads/966/QPP%20COVID-19%20Response%20Fact%20Sheet.pdf
[10] Refer to footnote 4.
[11] Refer to footnote 4.
[12] Nguyen, Kevin H., Kaitlyn P. Lew, and Amal N. Trivedi. "Trends in Collection of Disaggregated Asian American, Native Hawaiian, and Pacific Islander Data: Opportunities in Federal Health Surveys." *American Journal of Public Health* (2022).
[13] Kader, Farah, Lan N. Doan, Matthew Lee, Matthew K. Chin, Simona C. Kwon, and Stella S. Yi. "Disaggregating Race/Ethnicity Data Categories: Criticisms, Dangers, And Opposing Viewpoints", *Health Affairs Forefront* (2022).
[14] Centers for Medicare & Medicaid (CMS), Office of Minority Health. "Utilization of Z Codes for Social Determinants of Health among Medicare Fee-for-Service Beneficiaries." (2019) https://www.cms.gov/files/document/z-codes-data-highlight.pdf

Variable	Advantages	Disadvantages	Used in Testing
American Community Survey	• Can link beneficiary's ZIP code to socioeconomic (SES) measurement of their neighborhood • Many SES indices can be derived from the survey data (e.g., Agency for Healthcare Research and Quality SES index, deprivation index)	• Only a proxy measure, not always accurate at individual-level	Yes

First, this analysis evaluated each of the variables for their association with episode cost using step-wise regression. In these models, race is a categorical variable (reference = White race). Testing findings demonstrate that dual Medicare and Medicaid enrollment status is the most powerful predictor, even in the presence of other variables (Tables 9 and B1). This is also consistent with other research that found dual status to be the best proxy of SRFs in predicting health outcomes.[15]

Table 9: Associations of Available Social Risk Factor Variables and Cost of Care – TIN Reporting Level

Subgroup Risk Model	Variable	Coefficient (p-value)		
		Model 1: Base Model + Dual Status	Model 2: Base Model + Dual Status + Race	Model 3: Base Model + Dual Status + Race + AHRQ SES
PCI with NSTEMI Diagnosis	Dual Status	$578.64 (p<0.001)	$586.49 (p<0.001)	$556.37 (p<0.001)
	Race – Asian	-	$375.07 (p=0.09)	$437.02 (p=0.05)
	Race – Black	-	-$165.98 (p=0.18)	-$200.53 (p=0.11)
	Race – Hispanic	-	-$718.78 (p<0.001)	-$758.03 (p<0.001)
	Race – North American Native	-	$1,250.25 (p<0.001)	$1,197.43 (p<0.001)
	Race – Others	-	$145.29 (p=0.37)	$158.56 (p=0.33)
	Race – White	-	REF	REF
	AHRQ SES Index	-	-	-$14.37 (p=0.01)
PCI with STEMI Diagnosis	Dual Status	$1,149.31 (p<0.001)	$1,038.04 (p<0.001)	$875.63 (p<0.001)
	Race – Asian	-	$1,757.35 (p<0.001)	$1,957.44 (p<0.001)

[15]Office of the Assistant Secretary for Planning and Evaluation. "Second report to Congress on social risk and Medicare's value-based purchasing programs." (2020) https://aspe.hhs.gov/pdf-report/second-impact-report-to-congress

Subgroup Risk Model	Variable	Coefficient (p-value)		
		Model 1: Base Model + Dual Status	Model 2: Base Model + Dual Status + Race	Model 3: Base Model + Dual Status + Race + AHRQ SES
PCI with STEMI Diagnosis	Race – Black	-	$459.78 (p=0.03)	$301.18 (p=0.16)
	Race – Hispanic	-	-$108.19 (p=0.77)	-$315.22 (p=0.4)
	Race – North American Native	-	$1039.83 (p=0.12)	$853.95 (p=0.2)
	Race – Others	-	-$472.6 (p=0.06)	-$344.71 (p=0.17)
	Race – White	-	REF	REF
	AHRQ SES Index	-	-	-$67.48 (p<0.001)
PCI without NSTEMI or STEMI Diagnosis	Dual Status	$976.95 (p<0.001)	$841.84 (p<0.001)	$831.32 (p<0.001)
	Race – Asian	-	$737.91 (p=0.01)	$767.99 (p=0.01)
	Race – Black	-	-$457.2 (p<0.001)	-$456.03 (p<0.001)
	Race – Hispanic	-	$2,359.83 (p<0.001)	$2,355.3 (p<0.001)
	Race – North American Native	-	-$370.35 (p=0.47)	-$375.98 (p=0.47)
	Race – Others	-	-$101.24 (p=0.64)	-$92.52 (p=0.67)
	Race – White	-	REF	REF
	AHRQ SES Index	-	-	-$2.65 (p=0.72)

The subsequent analyses focus on dual status as the main proxy variable for SRFs for risk adjustment. To determine whether it's appropriate to risk adjust for SRFs, the following criteria are considered:

(i) whether there's an association between social risk and performance by examining the coefficient of patient-level dual status when added into the risk model,

(ii) whether the observed association is most influenced by patient-level factors or clinician-level factors by examining the stability of the patient-level dual status coefficient after adding clinician's dual share variable, as well as including the clinician's fixed effects,

(iii) whether the patient's need or complexity (rather than poor quality) is driving the observed performance differences by examining the differences in performance on dual patients versus non-dual patients and if there are many clinicians who are able to perform similarly or better on their dual patients than their non-dual patients, and

(iv) the impact of risk adjusting for SRFs by examining the performance shift of clinicians compared to a risk adjustment model that doesn't risk adjust for SRFs.

There's a statistically significant association between the patient's dual status and episode cost (Table 10). This association decreases but remains statistically significant after adding variables to account for provider-level factors, which suggests that the observed association with cost can

be partially driven by provider-level factors. There is not a large performance degradation with increasing share of dual episodes (Table 11). The vast majority of clinicians are still able to perform equally well or significantly better on their dual episodes than their non-dual episodes (Table 12). Lastly, risk adjusting for dual status appears to change the performance ranking for many clinicians, but the magnitude of change is small (Table 13).

Table 10: Coefficient of Patient-level Dual Status under Different Models (Revised Measure)

Level	Subgroup Risk Model	% of All Episodes	Coefficient of Patient-level Dual Status (P-value)		
			Base Model + Patient-level Dual Status	Base Model + Patient-level Dual Status + Clinician's Dual Share	Base Model + Patient-level Dual Status + Clinician's Fixed Effect
TIN	PCI with NSTEMI Diagnosis	45.99%	$578.64 (p<0.001)	$419.79 (p<0.001)	$424.25 (p<0.001)
TIN	PCI with STEMI Diagnosis	25.75%	$1,149.31 (p<0.001)	$819.01 (p<0.001)	$749.34 (p<0.001)
TIN	PCI without STEMI or NSTEMI Diagnosis	28.26%	$976.95 (p<0.001)	$595.38 (p<0.001)	$705.51 (p<0.001)
TIN-NPI	PCI with NSTEMI Diagnosis	47.00%	$557.13 (p<0.001)	$261.24 (p<0.001)	$375.98 (p<0.001)
TIN-NPI	PCI with STEMI Diagnosis	24.54%	$1,000.19 (p<0.001)	$382.57 (p=0.01)	$721.93 (p<0.001)
TIN-NPI	PCI without STEMI or NSTEMI Diagnosis	28.45%	$1,212.95 (p<0.001)	$511.98 (p<0.001)	$564.85 (p<0.001)

Table 11: Mean Ratio of Observed Cost to Expected Cost (O/E) Stratified by Clinician's Dual Share and Patient's Dual Status (Revised Measure)

Dual Share	TIN			TIN-NPI		
	All Episodes	Dual Episodes	Non-Dual Episodes	All Episodes	Dual Episodes	Non-Dual Episodes
All	0.99	1.01	0.99	0.99	1.01	0.99
0%	0.99	-	0.99	0.98	-	0.98
1-20%	0.99	1.01	0.99	0.99	1.01	0.99
21-40%	1.00	1.01	1.00	1.01	1.02	1.00
41-60%	1.01	1.01	1.00	1.00	1.00	1.00
61-80%	0.99	0.98	1.00	0.97	0.97	0.98
81-99%	0.93	0.94	0.84	-	-	-

Table 12: Proportions of Clinicians Who Perform Significantly Worse, Equally Well, or Significantly Better on Their Dual Episodes than Non-Dual Episodes (Revised Measure)

Reporting Level	Significantly Worse	Equally Well	Significantly Better
TIN	7.04%	92.61%	0.35%
TIN-NPI	6.46%	92.32%	1.21%

Table 13: Clinicians' Performance Shift Measured by the Change in the Average Ratio of Observed Cost to Expected Cost (O/E) (Revised Measure)

Reporting Level	Proportions of Clinicians Affected at Various Levels of Performance Shift	
	Ranking Shift by 1% or more	Ranking Shift by 5% or more
TIN	70.84%	5.64%
TIN-NPI	70.05%	4.68%

2.6 Impact of Exclusions

Table 14 displays descriptive statistics of all episodes meeting the revised measure's triggering logic, excluded episodes, and final reportable episodes at both TIN and TIN-NPI levels. These exclusion criteria ensure that the reportable episode populations are more homogenous and comparable than all episodes meeting triggering logic. It is worth noting that only the observed cost is shown, which has not been risk adjusted for using our risk adjustment model. Therefore, the differences in cost may appear much smaller after risk adjustment than as-is.

Overall, exclusion criteria decrease the observed cost of all episodes meeting trigger logic from the mean of $21,100 to $19,812 at the TIN-level and $20,149 at the TIN-NPI level. Almost all of the exclusion criteria have higher mean observed cost than all episodes meeting triggering logic. The largest exclusions come from applying the testing volume threshold to ensure a sufficient sample size for the measure.

Episodes where a beneficiary died before the episode end date are excluded because they do not provide sufficient data in the episode window period. These episodes also have a higher mean observed cost than all episodes meeting triggering logic, at $28,372, likely because the costs are distributed over fewer days than a typical episode.

Episodes classified as outlier cases are excluded because they deviate substantially from the projected cost for a given patient risk profile. Outlier episodes have a mean observed episode cost of $45,187 compared to $21,100 for all episodes meeting triggering logic. The wide variability of observed episode costs for outlier cases also supports their exclusion. At the 10th percentile the outlier cases observed cost is $18,526 and at the 90th percentile the observed cost is $77,633.

Episodes where there is not an attributed clinician are excluded because these episodes do not have any TIN-NPIs that billed at least 30% of the clinically-related claims with a relevant diagnosis**.** As such, they cannot be used in the measure at the TIN-NPI level.

Based on the input from the clinical expert workgroup, several subpopulations are excluded because these episodes can be clinically distinct from the overall population of patients receiving PCIs. Workgroup members recommended excluding patients with (i) history of intracranial hemorrhage or cerebral infarction, (ii) new cardiac device implantation, (iii) recent hospitalization for respiratory failure, (iv) recent hospitalization for STEMI, (v) shock, (vi) transplants, and (vii) ventilator dependence. These episodes have mean observed costs higher than all episodes meeting trigger logic, which suggests that they may have distinct resource use patterns from a typical major depression episode.

Table 14: Cost Statistics for Measure Exclusions (Revised Measure)

Exclusion Criteria	Episodes		Observed Episode Cost					
				Percentile				
	Count	Percent of All Episodes Meeting Trigger Logic	Mean	10th	25th	50th	75th	90th
All Episodes Meeting Triggering Logic	93,790	100.00%	$21,100	$13,935	$14,641	$18,340	$23,667	$31,638
Beneficiary Death in Episode	3,986	4.25%	$28,372	$17,318	$22,023	$24,099	$30,820	$43,495
Outlier Cases	1,510	1.61%	$45,187	$18,526	$21,656	$35,436	$65,515	$77,633
No Attributed TIN	1,605	1.71%	$30,174	$16,438	$22,558	$25,623	$34,755	$46,813
Not an IPPS Acute Hospital or Psychiatric Facility	79	0.08%	$24,611	$14,042	$15,698	$22,192	$28,354	$41,770
Overlapping IP Admission Stays	959	1.02%	$27,883	$18,790	$19,983	$25,066	$30,619	$39,601
History of Intracranial Hemorrhage or Cerebral Infarction	4,028	4.29%	$23,530	$14,019	$15,118	$21,971	$26,186	$36,579
Patients with New Cardiac Device Implantation	539	0.57%	$23,546	$13,916	$15,040	$21,461	$25,793	$36,723
Recent Hospitalization for Respiratory Failure	424	0.45%	$27,332	$15,447	$21,745	$23,482	$30,858	$39,920
Shock	813	0.87%	$29,818	$16,755	$21,664	$24,611	$33,701	$46,588
Recent Hospitalization for STEMI	8,756	9.34%	$22,258	$13,718	$14,696	$21,322	$24,704	$33,938
Transplant Patients	1,030	1.10%	$24,585	$14,152	$15,524	$22,226	$26,241	$37,197
Ventilator Dependence	348	0.37%	$25,622	$14,270	$20,030	$23,202	$28,012	$37,278
TIN does not Meet Testing Volume Threshold	11,231	11.97%	$21,010	$13,796	$14,539	$18,231	$23,551	$31,846

Exclusion Criteria	Episodes		Observed Episode Cost					
				Percentile				
	Count	Percent of All Episodes Meeting Trigger Logic	Mean	10th	25th	50th	75th	90th
TIN-NPI does not Meet Testing Volume Threshold	79,664	84.94%	$20,887	$13,920	$14,615	$17,942	$23,477	$31,091
Reportable Episodes (if all clinicians reported as TIN at the Testing Volume Threshold)	64,496	68.77%	$19,812	$13,925	$14,557	$16,450	$22,958	$29,087
Reportable Episodes (if all clinicians reported as TIN-NPI at the Testing Volume Threshold)	10,228	10.91%	$20,149	$13,932	$14,590	$17,014	$23,279	$30,008

Appendix A. Distributions of Measure Score (Revised Measure)

Figure 2: Distribution of Measure Score - TIN

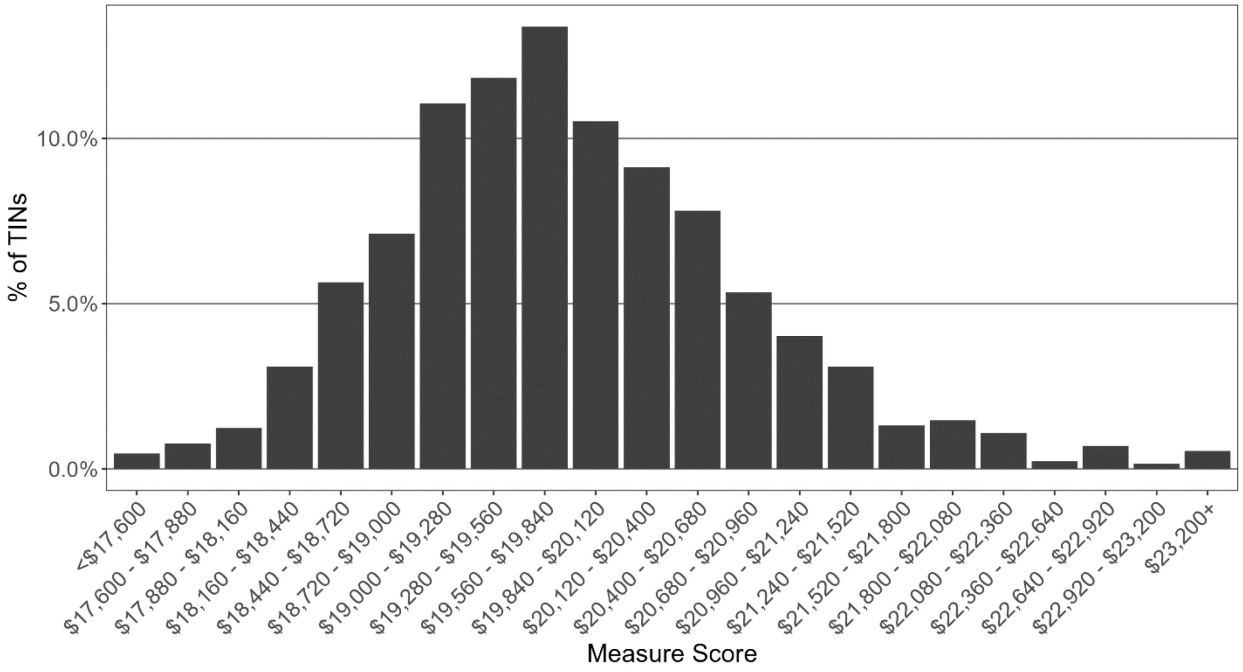
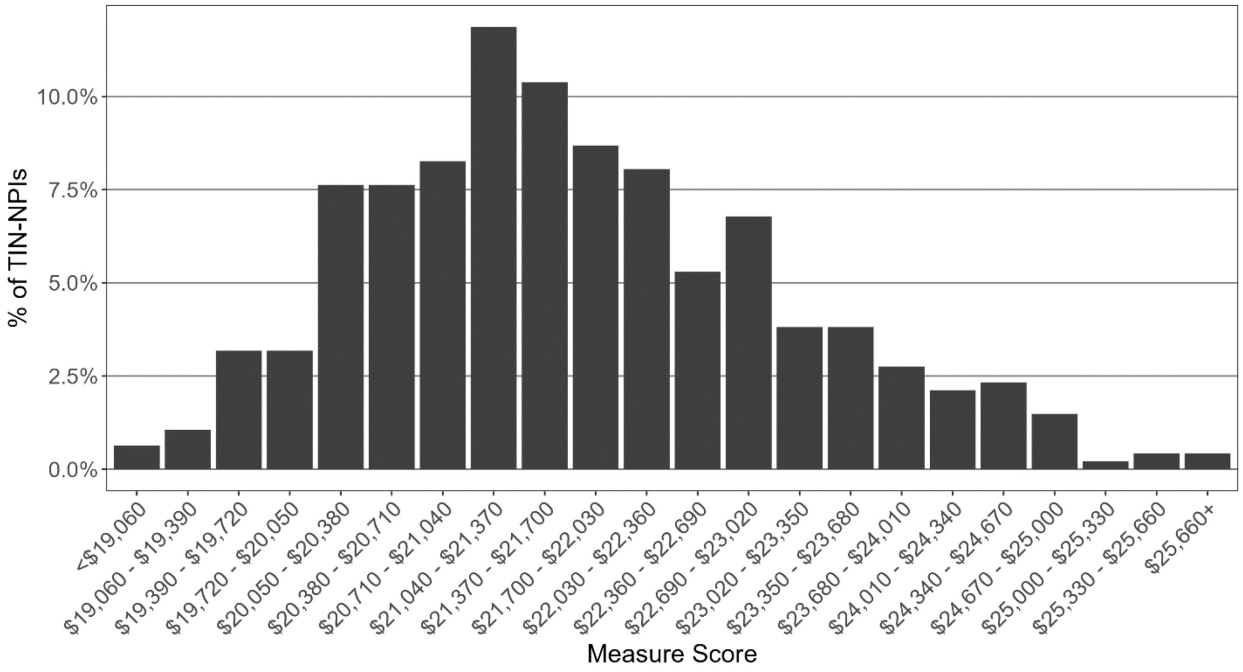


Figure 3: Distribution of Measure Score - TIN-NPI



Appendix B. Associations between Social Risk Factor Variables and Cost of Care for TIN-NPIs

Table B1. Associations of Available Social Risk Factor Variables and Cost of Care – TIN-NPI Reporting Level

Subgroup Risk Model	Variable	Coefficient (p-value)		
		Model 1: Base Model + Dual Status	Model 2: Base Model + Dual Status + Race	Model 3: Base Model + Dual Status + Race + AHRQ SES
PCI with NSTEMI Diagnosis	Dual Status	$557.13 (p<0.001)	$563.55 (p<0.001)	$531.31 (p<0.001)
	Race – Asian	-	$639.62 (p<0.001)	$713.01 (p<0.001)
	Race – Black	-	-$311.58 (p<0.001)	-$347.51 (p<0.001)
	Race – Hispanic	-	-$801.95 (p<0.001)	-$839.50 (p<0.001)
	Race – North American Native	-	$822.22 (p<0.001)	$763.02 (p=0.01)
	Race – Others	-	$607.33 (p<0.001)	$623.80 (p<0.001)
	Race – White	-	REF	REF
	AHRQ SES Index	-	-	-$15.41 (p<0.001)
PCI with STEMI Diagnosis	Dual Status	$1,000.19 (p<0.001)	$910.14 (p<0.001)	$768.73 (p<0.001)
	Race – Asian	-	$3,132.14 (p<0.001)	$3,354.48 (p<0.001)
	Race – Black	-	-$290.51 (p=0.11)	-$445.21 (p=0.01)
	Race – Hispanic	-	-$621.76 (p=0.06)	-$817.85 (p=0.01)
	Race – North American Native	-	$292.45 (p=0.61)	$88.25 (p=0.88)
	Race – Others	-	-$151.85 (p=0.50)	-$28.96 (p=0.90)
	Race – White	-	REF	REF
	AHRQ SES Index	-	-	-$62.58 (p<0.001)
PCI without NSTEMI or STEMI Diagnosis	Dual Status	$1,212.95 (p<0.001)	$1,066.96 (p<0.001)	$1,089.99 (p<0.001)
	Race – Asian	-	$785.92 (p<0.001)	$791.00 (p<0.001)
	Race – Black	-	-$376.71 (p<0.001)	-$335.41 (p=0.01)
	Race – Hispanic	-	$3,480.95 (p<0.001)	$3,502.56 (p<0.001)
	Race – North American Native	-	-$752.83 (p=0.06)	-$712.11 (p=0.07)

Subgroup Risk Model	Variable	Coefficient (p-value)		
		Model 1: Base Model + Dual Status	Model 2: Base Model + Dual Status + Race	Model 3: Base Model + Dual Status + Race + AHRQ SES
PCI without NSTEMI or STEMI Diagnosis	Race – Others	-	-$605.02 (p<0.001)	-$617.81 (p<0.001)
	Race – White	-	REF	REF
	AHRQ SES Index	-	-	$13.31 (p=0.03)