Quality Payment PROGRAM

Low Back Pain

Measure Testing Form

January 2022

Contents

1.0 Introduction

This Measure Testing Form (MTF) provides a brief summary of the preliminary measure testing results as part of field testing five episode-based cost measures. Stakeholders may review these results, alongside other documentation, to provide feedback on the draft measure using the field testing survey. The testing results reflect the performance of the measure as specified at the time of field testing, which is part of the measure development process. 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 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 “Physician Cost Measures and Patient Relationship Codes (PCMP).” The contract number is 75FCMC18D0015, Task Order 75FCMC19F0004.

1.2 Measure Name

Low Back Pain

1.3 Type of Measure

Cost/Resource Use

1.4 Data

The study period is January 1, 2019 through December 31, 2019. 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, B, and D administrative claims data, Long-Term Minimum Data Set, and the Medicare Enrollment Database. For testing purpose, other data sources are used, including the American Community Survey, Common Medicare Environment, and Uniform Data System.

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 clinician 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 measure as specified for field testing. Section 2.1 provides an overview of the measure's coverage of beneficiaries and cost. Section 2.2 lists the most frequently attributed specialties. Sections 2.3 through 2.5 provide evidence of scientific acceptability of the measure. Section 2.6 presents empirical results of the risk adjustment and stratification methods used by this measure. Section 2.7 examines the impact of adding social risk factors to the measure's risk adjustment model. Lastly, Section 2.8 examines the impact of exclusion criteria used by the measure through their frequency and resource use patterns.

2.1 Measure Coverage

Table 1 shows the number of patients and the amount of Medicare Parts A and B cost covered by this measure. Given the high prevalence of low back pain, this measure has the potential to impact care provided to a large number of patients and to influence Medicare spending to promote higher value care. Table 1 shows the Low Back Pain measure captures approximately 4 million patients, and about 3 percent of all Parts A and B costs.

Table 1. Beneficiary and Cost Coverage

Coverage Metrics	Coverage at Volume Thresholds	
	1 Episode	20 Episodes
Number of Patients	4,128,199	3,872,740
Percent of Parts A & B Costs – TIN	3.0%	2.9%
Percent of Parts A & B Costs – TIN-NPI	3.0%	2.4%

2.2 Frequently Attributed Specialties

Table 2 shows the top 10 attributed specialties for this measure, using a 20-episode episode testing volume threshold. The most frequently observed attributed specialties align with the intent of the Low Back Pain measure, which is to capture the role of clinicians who provide outpatient care and management of low back pain, including specialists and primary care clinicians. These specialties are also consistent with input provided by stakeholders, including patient and family partners (PFPs), during the measure development process. PFPs identified physical therapists, chiropractors, primary care clinicians, physiatrists, orthopedic surgeons, and pain specialists amongst others as being part of their care team. The Draft Cost Measure Methodology contains more information about the measure's attribution methodology.

Table 2 also provides metrics to show the breakdown of the share of episodes covered by each specialty. Chiropractors make up the largest share of attributed clinicians who meet the testing volume threshold (34.7%). Physical therapists are the second most frequently attributed specialty, comprising 15.8% of all attributed clinicians.

Finally, Table 2 shows the percentage of each specialty that is covered by this measure. More than half of all chiropractors (54.8%) who billed at least one Part B physician/supplier claim in 2019 have at least 20 Low Back Pain episodes, reflecting the high prevalence of low back pain. One note about this metric is that specialties vary in size; for instance, the share of internal medicine clinicians covered by the measure appears low (4.2%) but it is a very large specialty.

Table 2. Count and Attribution Frequency of the Top 10 Attributed Specialties at a Testing Volume of 20 Episodes

Rank	Specialty	Number of TIN-NPIs Attributed	Percent of Specialty Among All Attributed TIN-NPIs	Percent of All Episodes	Percent of Specialty (TIN-NPI) Attributed to Measure
1	Chiropractic	24,200	34.7%	35.6%	54.8%
2	Physical Therapist in Private Practice	11,046	15.8%	9.6%	14.7%
3	Internal Medicine	5,803	8.3%	5.7%	4.2%
4	Family Practice	5,381	7.7%	5.1%	4.7%
5	Orthopedic Surgery	3,619	5.2%	5.7%	12.9%
6	Physical Medicine and Rehabilitation	3,524	5.1%	8.0%	33.3%
7	Nurse Practitioner	3,020	4.3%	4.1%	1.6%
8	Anesthesiology	2,396	3.4%	5.3%	4.4%
9	Pain Management	2,294	3.3%	5.7%	62.7%
10	Physician Assistant	2,142	3.1%	3.1%	1.8%

2.3 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 3 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 field testing, we used a 20-episode threshold (bolded in the table below); for simplicity, we use this threshold across all measures. If the measure is implemented in the Merit-based Incentive Payment System (MIPS) in the future, CMS will establish a case minimum through notice-and-comment rulemaking.

At the testing volume of 20 episodes, the mean reliability for TINs is 0.882 and for TIN-NPIs is 0.829 (Table 3). CMS generally considers 0.4 as the threshold indicating 'moderate' reliability, which is supported by previous work into reliability and the threshold was finalized in the CY 2022 Physician Fee Schedule final rule.[2,3] Almost all of TINs and TIN-NPIs meet or exceed 0.4 reliability at the 20-episode testing volume threshold, with 99.90% of TINs and 99.46% of TIN-NPIs meeting or exceeding the 0.4 threshold.

[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.

Table 3. Sample Size, Mean Reliability, and Proportion of Clinicians above Moderate Reliability at Various Testing Volume Thresholds

Testing Volume Threshold	TIN			TIN-NPI		
	Number of TINs	Mean Reliability	Percent Above 0.4	Number TIN-NPIs	Mean Reliability	Percent Above 0.4
10	68,112	0.83	97.8%	125,523	0.75	93.3%
20	**49,949**	**0.88**	**99.9%**	**69,742**	**0.83**	**99.5%**
30	38,740	0.91	100.0%	46,286	0.87	100.0%

2.4 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 performance of clinicians. Validity is tested empirically by examining the association between the measure score and high-cost events that drive the measure score, such as downstream complications and consequences of care, and the correlation between cost and quality measures.

The measure score reflects the average risk-adjusted and specialty-adjusted cost of episodes that were attributed to a particular clinician or group. The risk-adjusted and specialty-adjusted cost neutralizes the effects of risk factors deemed to be outside of a clinician's influence (e.g., pre-existing conditions, age, or indicators of clinical severity) on the standardized cost[4] of clinically relevant services observed during a care episode.

The Low Back Pain measure stratifies episodes into four mutually exclusive and exhaustive sub-groups based on whether the episode includes spinal surgery and whether the patient has complex low back pain. The risk adjustment model is run separately within each sub-group, which means that episodes are only compared with like episodes in the same sub-group. For more information about risk adjustment, specialty adjustment, and sub-groups for this measure, please see the Draft Cost Measure Methodology.

The testing results confirm that the cost differential between episodes with and without spine surgery is neutralized so that providers are not disadvantaged for having episodes with spine surgery or treating patients with complex low back pain. Specifically, the mean observed over expected cost ratio (O/E) are close to 1.00 for providers with at least one episode classified in each of the four sub-groups. In other words, on average, the actual cost is very close to the projected cost because episodes are only benchmarked against others in the same subgroup and not across subgroup.

When comparing TINs with episodes where the patient has complex low back pain, the results show that whether or not there was a spine surgery does not adversely impact scores:

- 49,806 TINs have at least one episode in the sub-group "no spine surgery and the patient has complex low back pain". The mean O/E for these TINs is 0.97.
- 16,640 TINs have at least one episode in the sub-group "spine surgery with complex low back pain". The mean O/E for these TINs is 0.98.

[4] Claim payments are standardized to account for differences in Medicare payments for the same service(s) across Medicare providers. Payment standardized costs remove the effect of differences in Medicare payment among health-care providers that are the result of differences in regional health-care provider expenses measured by hospital wage indexes and geographic price cost indexes (GPCIs) or other payment adjustments, such as those for teaching hospitals. For more information, please refer to the "CMS Price (Payment) Standardization - Basics" and "CMS Price (Payment) Standardization - Detailed Methods" documents posted on the Payment Standardization webpage. (https://resdac.org/articles/cms-price-payment-standardization-overview)

The results are similar when comparing TINs with episodes where the patient does not have complex low back pain:

- 49,935 TINs have at least one episode in the sub-group “no spine surgery without complex low back pain”. The mean O/E for these TINs is 1.02.
- 9,421 TINs have at least one episode in the sub-group “spine surgery without complex low back pain”. The mean O/E for these TINs is 1.01.

Table 4 shows a subset of potentially high-costs events that could influence performance on the Low Back Pain cost measure, as well as the mean score associated with how frequently each type of high-cost events occurs. Across all events shown, providers with low frequencies of high cost events had minimal or no increase in mean score compared to the mean score for all final episodes. As expected, higher frequencies of high-costs events are associated with slight increases in mean scores. The increases in mean scores are most pronounced among TINs and TIN-NPIs with higher frequencies of imaging.

Table 4: Mean and Standard Deviation of Score Stratified by the Frequency of Observing High-Cost Events

High-Cost Events	TIN			TIN-NPI		
	Quartiles (Frequency Range)	Mean Score	Score Standard Deviation	Quartiles (Frequency Range)	Mean Score	Score Standard Deviation
All Final Episodes	All	$1,876	$570	All	$1,935	$598
Acute Inpatient Stay	0%	$1,842	$596	0%	$1,897	$618
	Q1: 0.1% - 1.5%	$1,843	$502	Q1 (0.2% - 2%)	$1,879	$525
	Q2: 1.5% - 2.5%	$1,874	$499	Q2 (2% - 3.3%)	$1,955	$556
	Q3: 2.5% - 4%	$1,923	$545	Q3 (3.3% - 5%)	$2,012	$607
	Q4: 4% - 71.9%	$2,034	$557	Q4 (5% - 88.5%)	$2,094	$557
Post-Acute Care	0%	$1,768	$545	0%	$1,828	$576
	Q1 (0.2% - 2.5%)	$1,832	$494	Q1 (0.2% - 3.1%)	$1,881	$512
	Q2 (2.5% - 4.5%)	$1,884	$548	Q2 (3.1% - 5.1%)	$1,948	$600
	Q3 (4.5% - 8.6%)	$1,876	$543	Q3 (5.1% - 9.5%)	$1,939	$561
	Q4 (8.6% - 96.7%)	$2,087	$661	Q4 (9.5% - 97%)	$2,160	$684
Imaging	0%	$1,454	$429	0%	$1,392	$404
	Q1 (1.7% - 27%)	$1,734	$543	Q1 (1.7% - 28.6%)	$1,774	$579
	Q2 (27% - 35.9%)	$1,866	$569	Q2 (28.6% - 40%)	$1,929	$603
	Q3 (35.9% - 50%)	$1,891	$577	Q3 (40% - 57.4%)	$1,948	$593
	Q4 (50% - 100%)	$2,016	$553	Q4 (57.4% - 100%)	$2,091	$575
Spine Injections	0%	$1,662	$652	0%	$1,717	$710
	Q1 (0.5% - 6.3%)	$1,770	$571	Q1 (0.5% - 7.4%)	$1,801	$602
	Q2 (6.3% - 11.1%)	$1,862	$564	Q2 (7.4% - 14.3%)	$1,919	$597

High-Cost Events	TIN			TIN-NPI		
	Quartiles (Frequency Range)	Mean Score	Score Standard Deviation	Quartiles (Frequency Range)	Mean Score	Score Standard Deviation
	Q3 (11.1% - 19.6%)	$1,891	$550	Q3 (14.3% - 30%)	$1,980	$565
	Q4 (19.6% - 100%)	$2,027	$535	Q4 (30% - 100%)	$2,082	$561

We also examined the correlation between the cost measure and related quality measures to test the relationship between these different metrics. While there are important limitations to this analysis, it can provide some indication of whether there is variation in cost with different levels of quality. In brief, we note the following key points for interpreting the strength and direction of correlations:

- A strong inverse correlation – good performance on cost with poor quality performance – would indicate that variation in cost is solely reflective of variation in quality. This suggests that care stinting could be a concern.
- A weak correlation between cost and quality in either a positive or negative direction indicates variation in cost at any given level of quality. This suggests that cost performance can be improved without negatively impacting quality.

In general, the direction of correlations indicate the following:

- Positive correlations with quality measures indicate that clinicians providing better quality care on that particular metric tend to also have lower costs. That is, clinicians who have high rates of performing specific quality actions (as measured through process measures) or achieve better patient health outcomes (as measured through outcome measures) tend to have lower costs of care. As such, these associations could represent ways to lower costs while also providing high-quality care.
- A negative correlation between a cost and quality measure does not indicate an absence of cost improvement potential consistent with high-value care. This is because other approaches to improving cost performance (e.g., patient education) may not be captured by the selected quality measure.

There are several key considerations regarding the conceptual relationship between measures and data limitations when interpreting the results. The extent to which correlations can provide meaningful information depends on what is being measured. Ideally, measures should apply to the same care provided for the same patient cohort for the same time horizon. Correlations with a quality measure that focuses on outcomes for the same patient cohort may be more informative than a broadly applicable process measure that applies across a wide range of conditions. Clinicians select only 6 MIPS quality measures to report and are generally required to only report one outcome or high-priority measure. This selective reporting likely biases the observed sample.

Table 5 shows the correlation between the Low Back Pain cost measure and related MIPS quality measures. To aid interpretability, non-inverse quality measure scores were inverted so that a lower score indicates better quality performance, which is the same direction as the cost measure. Inverse quality measure scores were not adjusted. Because of this, correlations can be interpreted the same way, regardless of whether a quality measure is inverse or non-inverse. Quality measures are selected based on their clinical proximity to the cost measure, such as assessing quality actions related to a similar patient cohort, and the number of clinicians with both cost and quality measures.

There are very low numbers of TINs and TIN-NPIs that have both the quality and cost measure, providing a limited sample for this analysis. There are 49,949 TINs and 69,742 TIN-NPIs that meet the testing volume threshold for Low Back Pain; across all the quality measure pairs tested, the highest number of overlapping providers was 2,158. In general, the results are weakly correlated in either direction.

The exception is one of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) for MIPS measures; there is a statistically significant strong positive correlation between performance on the Low Back Pain cost measures and CAHPS7: CAHPS for MIPS SSM (Summary Survey Measure): Health Status and Functional Status. The CAHPS 7 measure is only reported at the TIN level, and 57 of the TINs with the Low Back Pain measure reported CAHPS 7. The strong positive correlation between CAHPS 7 and the Low Back Pain measure suggests that TINs with higher quality scores on CAHPS 7 tend to also perform better on the Low Back Pain measure (i.e., they have lower costs). However, it is important to note the limitations due to the very small sample size.

Table 5. Correlation between Quality Measures and Cost Measure

Related Quality Measure (Type)	TIN or TIN-NPI	Number of Entities with Both Cost and Quality Measures Available	Pearson Correlation	P-value
Q182: Functional Outcome Assessment (Process)	TIN	817	-0.05	0.147
	TIN-NPI	2,158	-0.06	0.003*
Q220: Functional Status Change for Patients with Low Back Impairments (Patient-Reported Outcome)	TIN	58	-0.17	0.213
	TIN-NPI	208	0.00	0.992
Q223: Functional Status Change for Patients with General Orthopedic Impairments (Patient-Reported Outcome)	TIN	56	-0.20	0.132
	TIN-NPI	193	-0.04	0.582
CAHPS7: CAHPS for MIPS SSM: Health Status and Functional Status (Patient-Reported Outcome)	TIN	57	0.49	0.000*

*P-value <0.05 indicates statistical significance

2.5 Performance Gap

Table 6 shows the distribution of the measure score for TINs and TIN-NPIs. The score interquartile range (IQR) for both TINs and TIN-NPIs is greater than 30 percent of the mean score. Additionally, for both TINs and TIN-NPIs, the 90th percentile score was more than twice the 10th percentile score. The distributions show meaningful variation in cost performance and suggest that there is room for improvement in the costs of care for a low back pain episode.

Table 6. Distribution of the Measure Score

Metric	TIN	TIN-NPI
Mean Score	$1,876	$1,935
Score Interquartile Range (IQR)	$643	$690
Standard Deviation	$570	$598
Coefficient of Variation	0.30	0.31
Score Percentile		
10th	$1,258	$1,285
25th	$1,507	$1,539
50th	$1,802	$1,853
75th	$2,150	$2,229
90th	$2,562	$2,659

2.6 Risk Adjustment and Stratification

2.6.1 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 for the measure is 0.638, and 0.638 after adjusting for the model's complexity based on the number of risk adjustors used. In other words, 63.8% 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.

2.6.2 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 1.0 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 risk adjustment model is consistent, with the average predictive ratios observed to be close to 1.00 across all deciles, with the range between 0.99 and 1.01. Overall, the risk adjustment model does not over- or under-predict cost across the full range of resource use patterns in the population.

Table 7. Predictive Ratio by Decile of Predicted Episode Cost

Decile	Average Predictive Ratio
Decile 1	1.00
Decile 2	0.99
Decile 3	0.99
Decile 4	1.01
Decile 5	1.01
Decile 6	0.99
Decile 7	1.00
Decile 8	1.01
Decile 9	1.01
Decile 10	1.00

2.7 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.[5]

To examine the extent that SRFs can mask the underlying performance of clinicians, this analysis added SRFs to the base risk adjustment model to examine their potential impact to clinicians' scores and goodness of fit of the risk adjustment model. The base risk adjustment model includes the standard set of risk adjustors from the CMS-Hierarchical Condition Categories (HCC) version 22 in 2016, disability status, End-Stage Renal Disease (ESRD) status, comorbidity interaction variables, recent long-term care use, HCC count, and measure-specific clinical risk adjustors. For the full list of factors that were risk adjusted for this measure, please see the Draft Cost Measure Methodology.

The base model was compared against 3 models that included additional SRFs from the American Community Survey, Common Medicare Environment described below:

1. Dual Medicare-Medicaid eligibility status:
    - Base model
    - Medicare-Medicaid eligibility status: full, partial, or non-dual status
2. All Socioeconomic Status (SES) Variables Model:
    - Base model
    - Dual Medicare-Medicaid eligibility status: full, partial, or non-dual status

[5] https://qpp-cm-prod-content.s3.amazonaws.com/uploads/966/QPP%20COVID-19%20Response%20Fact%20Sheet.pdf

- Sex: female or male
- Race: Asian, Black, Hispanic, North American Native, White, and other
- Neighborhood Income Level: low income (median income < 33rd percentile nationally), medium income (median income between 33rd and 66th percentiles), and high income (median income > 66th percentile)
- Neighborhood Education Rate: majority less than high school, high school, or greater than high school
- Neighborhood Employment Rate: less than or greater than 10%

3. Agency for Healthcare Research and Quality SES Index Model:
    - Base model
    - Dual Medicare-Medicaid eligibility status
    - Sex: female or male
    - Race: Asian, Black, Hispanic, North American Native, White, and other
    - AHRQ SES Index Score: calculated using the AHRQ's scoring algorithm, based on data from the American Community Survey

Table 8 shows the percent of providers who would see their performance shifted, based on their percentile ranking, using the new SRF models. There are marginal changes to the performance at both TIN and TIN-NPI reporting levels compared to the base model. Across the 3 models, over 98% TINs and TIN-NPIs would experience small and likely inconsequential swing in their performance, of less than 5% compared to the base model. Additionally, the performance ranking has Spearman correlation values of greater than 0.99 between the base model and each of the SRF models, which also suggests that adjusting for SRF variables does not fundamentally change how the measure evaluates provider performance

Table 8. Clinicians' Performance Shift Measured by the Change in the Average Ratio of Observed-to-Expected Cost

Model	Reporting Level	Proportion of Clinicians Affected at Various Levels of Performance Shift						
		-10% or more	-9% to -6%	-5% to -2%	-1% to 1%	2% to 5%	6% to 9%	10% or more
Model 1	TIN	*	0.66%	5.06%	93.74%	0.54%	*	*
	TIN-NPI	*	0.59%	4.46%	94.31%	0.63%	*	*
Model 2	TIN	0.02%	0.66%	17.17%	63.78%	18.20%	0.17%	*
	TIN-NPI	0.02%	0.56%	16.10%	66.56%	16.59%	0.18%	*
Model 3	TIN	0.03%	1.13%	16.97%	62.42%	19.23%	0.22%	*
	TIN-NPI	0.03%	0.99%	15.87%	65.27%	17.60%	0.23%	*

Note: Cells with stars ('*') indicate values that were suppressed as there were less than 10 observations.

2.8 Impact of Exclusions

Table 9 displays descriptive statistics of all episodes meeting the 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.

The mean observed episode costs for reportable episodes at the TIN and TIN-NPI level ($1,940 and $1,805, respectively) are slightly lower than the mean observed episode cost for all episodes meeting the triggering logic ($2,103). All episode exclusion criteria shown below had higher mean observed episode costs compared to all episodes meeting triggering logic. All episode-based cost measures excluded outlier cases, which is supported for the Low Back Pain measure findings; the mean observed cost for these episodes was more than twice as costly as the mean observed cost of reportable episodes. Measure specific exclusion criteria (e.g., cauda equine syndrome, osteoporotic compression fracture, spinal infection, myelopathy, and trauma) also tended to have higher mean observed costs.

Table 9: Cost Statistics for Measure Exclusions

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	6,384,089	100.00%	$2,103	$162	$361	$875	$1,879	$4,071
Episode Length Less Than 120 Days	48,417	0.76%	$5,184	$469	$910	$2,012	$4,497	$9,214
No Attributed Clinician (TIN-NPI Reporting Only)	400,687	6.28%	$3,040	$465	$789	$1,416	$2,664	$6,212
Outlier Cases	120,427	1.89%	$5,354	$152	$309	$6,556	$7,236	$11,410
Cauda Equina syndrome	6,361	0.10%	$5,318	$400	$901	$2,030	$4,896	$11,150
Osteoporotic Compression Fracture	28,879	0.45%	$4,506	$346	$839	$1,987	$4,882	$9,695
Spinal Infection	18,985	0.30%	$5,673	$386	$890	$2,000	$4,790	$10,897
Myelopathy	127,569	2.00%	$3,563	$233	$580	$1,308	$2,765	$6,951
Trauma	105,463	1.65%	$4,326	$342	$790	$1,797	$4,405	$9,056
Reportable Episodes (if all clinicians reported as TIN at the testing volume threshold)	5,483,009	85.89%	$1,940	$158	$357	$856	$1,788	$3,595
Reportable Episodes (if all clinicians reported as TIN-NPI at the testing volume threshold)	3,961,881	62.06%	$1,805	$133	$311	$778	$1,680	$3,341