Quality Payment
PROGRAM

Low Back Pain

Measure Justification Form

September 2022

Table of Contents

1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the Low Back Pain measure. The form 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

Physician Cost Measure and Patient Relationship Codes

1.2 Date

Information included is current on September 27, 2022.

1.3 Project 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 Measure and Patient Relationship Codes (PCMP)." The contract number is 75FCMC18D0015, Task Order 75FCMC19F0004.

1.4 Measure Name

Low Back Pain Episode-Based Cost Measure

1.5 Type of Measure

Cost/Resource Use

1.6 Measure Description

The Low Back Pain episode-based cost measure evaluates a clinician's or clinician group's risk-adjusted and specialty-adjusted cost to Medicare for patients receiving care to manage and treat low back pain. This chronic condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during a Low Back Pain episode.

[1]CMS, "Low Back Pain Measure Methodology" and "Low Back Pain Measure Codes List" *MACRA Feedback Page*, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback

2.0 Importance

2.1 Evidence to Support the Measure Focus

The Low Back Pain measure was developed for use in the Merit-based Incentive Payment System (MIPS) to meet the requirements of the Social Security Act section 1848(r), added by MACRA. MIPS aims to reward high-value care by measuring clinician performance through 4 areas:

- quality
- improvement activities
- Promoting Interoperability
- cost

Each category assesses different aspects of care, and the categories are weighted such that they're combined into one composite score. CMS is introducing MIPS Value Pathways (MVPs) as a way to align and connect quality measures, cost measures, and improvement activities across performance categories of MIPS for different specialties or conditions. MVPs aim to provide a holistic assessment of clinician value for a specific type of care to achieve better healthcare outcomes and lower costs for patients.

The use of cost measures is required by statute, and their purpose is to assess resource use. To be effective, they should capture costs related to a clinician's care decisions and account for factors outside of their influence. This measure provides clinicians with information about their costs of care that they can use to understand the costs associated with their decision-making. Clinicians play an important role in healthcare expenditures' variation due to their ability to affect costs[2]. A cost measure offers opportunity for improvement if clinicians can exercise influence on the intensity or frequency of a significant share of costs during the episode, or if clinicians can achieve lower spending and better quality of care quality through changes in clinical practice.

The Low Back Pain episode-based cost measure was recommended for development through feedback gathered during a public comment period. The public recommended this measure because of the prevalence of low back pain and the variable costs associated with the management of the disease and its complications, as well as the chronic nature of condition. A measure-specific Clinician Expert Workgroup was then convened with clinicians, health care experts, and patient representatives who have appropriate experience to provide extensive, detailed input on this measure throughout its development.

Low back pain is one of the leading sources of years lived with disability.[3] Low back pain is increasingly common in the US, with roughly 20 percent of Americans experiencing low back

[2]David Cutler et al., "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," American Economic Journal: Economic Policy 11, no. 1 (February 1, 2019): 192–221, https://doi.org/10.1257/pol.20150421.

[3] A., Carter, A., Zipkin, B., Sartorius, B., Serdar, B., Sykes, B. L., Troeger, C., Fitzmaurice, C., … Murray, C. (2018). The State of US Health, 1990-2016: Burden of Diseases, Injuries, and Risk Factors Among US States. JAMA, 319(14), 1444–1472. https://doi.org/10.1001/jama.2018.0158

pain each year,[4,5] and about 6 percent of Americans requiring ambulatory visits as a result of this condition.[6] Prevalence is much higher among seniors: one 2008 study of 522 community-dwelling seniors in the mid-Atlantic found a low back pain prevalence of 48 percent of participants, nearly half of all who sought care.[7] Larger studies corroborate these findings: a 1985 study of 3,097 rural Iowans aged 65 years and older found that low back pain was reported by 23.6 percent of women and 18.4 percent of men who were surveyed.[8]

Low back pain prevalence has continued to grow; between 1991 and 2002, patients seeking low back treatment through Medicare grew at nearly triple the rate of Medicare beneficiary growth (131 percent versus 42 percent).[9] A study surveyed a representative sample of North Carolina households found that obesity, depression, an aging population, and awareness/willingness to seek treatment are linked to increases in low back pain across all subpopulations of age, race, and gender.[10] In addition, the growing prevalence of low back pain may result in other risk factors including opioid abuse since a majority (54.9%) of adults over 65 with an opioid diagnosis also had a diagnosis of low back pain.[11]

Health care expenditures attributable directly to low back pain were estimated at $23 billion in 1998, with a total cost of treating patients with low back pain of $90 billion. In 1998, these patients incurred a higher per capita cost than patients without low back pain, at $3,500 versus $2,100 per capita.[12] These numbers continue to grow: expenditures on spine treatments increased 65 percent between 1998 and 2008, up to $86 billion per year.[13] A more recent study found that low back and neck pain contributed the most to US health care spending among 154 mutually exclusive diagnoses, at $134.5 billion in 2016.[14]

Despite these rising costs, there has been little improvement in patient outcomes. A 2009 study documents increased spending on narcotics and epidural steroid injections between 1998 and 2008 at 423% to treat low back pain, which “haven’t been accompanied by population-level

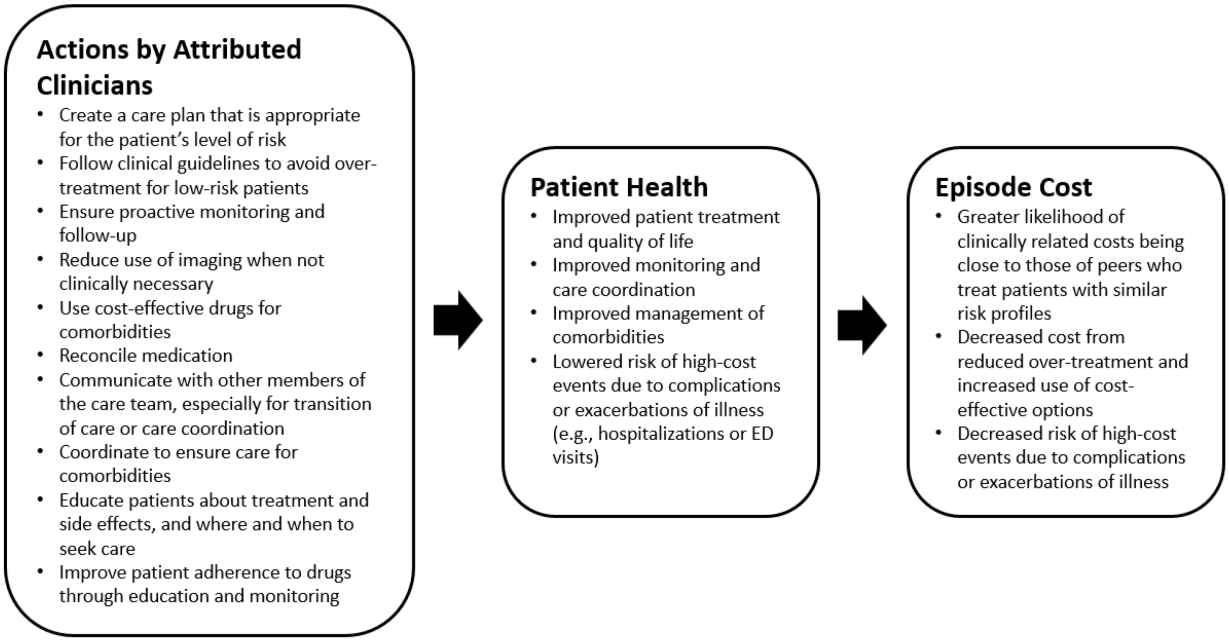
---

[4] Will, Joshua Scott, David Bury, and John Miller. “Mechanical Low Back Pain.” American Academy of Family Physicians 98(7) (2018): 421-428.
[5] Blanpied et al. “Neck Pain: Revision 2017: Clinical Practice Guidelines Linked to the International Classification of Functioning, Disability and Health From the Orthopaedic Section of the American Physical Therapy Association.” Journal of Orthopaedic & Sports Physical Therapy 47(7) (2017): A1-A83. doi:10.2519/jospt.2017.0302
[6] Davis, Matthew. “Where the United States Spends its Spine Dollars: Expenditures on different ambulatory services for the management of back and neck conditions.” Spine 37(19) (September 1 2012): doi:10.1097/brs.0B013E3182541F45.
[7] Hicks, Gregory, Jean Gaines, Michelle Shardell, and Eleanor Simonsick. “Associations of back and leg pain with health status and functional capacity of older adults: Findings from the retirement community back pain study.” Arthritis Care & Research 59(9) (29 August 2008): doi:10.1002/art.24006.
[8] Lavsky-Shulan M, Wallace RB, Kohout FJ, Lemke JH, Morris MC, Smith IM. Prevalence and functional correlates of low back pain in the elderly: the Iowa 65+ Rural Health Study. J Am Geriatr Soc. 1985 Jan;33(1):23-8. doi: 10.1111/j.1532-5415.1985.tb02855.x. PMID: 3155530.
[9] Weiner, Debra, Young-Sin Kim, Paula Bonino, and Tracy Wang, “Low Back Pain in Older Adults: Are We Utilizing Healthcare Resources Wisely?” Pain Medicine 7(2) (2006): 143-150. doi:10.1111/j.1526-4637.2006.00112.
10 Freburger et al. “The Rising Prevalence of Chronic Low Back Pain” in Arch Intern Med, 169(3) (2009): 251-258. doi:10.1001/archinternmed.2008.543.
[11] Hogans BB, Siaton BC, Taylor MN, Katzel LI, Sorkin JD. Low Back Pain and Substance Use: Diagnostic and Administrative Coding for Opioid Use and Dependence Increased in U.S. Older Adults with Low Back Pain. Pain Med. 2021 Apr 20;22(4):836-847. doi: 10.1093/pm/pnaa428. PMID: 33594426; PMCID: PMC8599750.
[12] Luo, Xuemei, Ricardo Pietrobon, Shawn Sun, Gordon Liu, and Lloyd Hey. “Estimates and Patterns of Direct Health Care Expenditures Among Individuals With Back Pain in the United States.” Spine 29(1) (2004): 79-86. doi:10.1097/01.BRS.0000105527.13866.0
[13] Davis, Matthew. “Where the United States Spends its Spine Dollars: Expenditures on different ambulatory services for the management of back and neck conditions.” Spine 37(19) (September 1 2012): doi:10.1097/brs.0B013E3182541F45.
[14] Dieleman, Joseph, Jackie Cao, and Abby Chapin. “US Health Care Spending by Payer and Health Condition, 1996-2016.” JAMA Network 323(9) (2020): 863-884. doi:10.1001/jama.2020.0734.

improvements in patient outcomes or disability rates."[15] These increases in resource use, despite modest increases in condition prevalence and a lack of improvements in patient outcomes, suggest the need for more precise measures of resource use and quality of care.

2.1.1 Logic Model

Figure 1. Logic Model of Steps between Actions by Attributed Clinicians and Episode Cost



2.2 Performance Gap

2.2.1 Rationale

According to the literature and feedback received through stakeholder input activities, the Low Back Pain measure's focus represents an area where there are substantial opportunities for improvement, namely, reducing low back pain complications through early physical therapy/conservative care, and promoting cost efficiency by avoiding expensive treatment options (e.g., imaging, injections) during the initial stages of care.

Research indicates that early care choices are important to downstream costs. Numerous studies have analyzed the relationship between receipt of physical therapy (PT) during early stages of care (e.g., within 14 days of pain consultation), and the incidence of downstream low back pain complications and health expenditures. For example, studies by Fritz et al (2008, 2013, 2015) have found that early, guideline-adherent PT is associated with both significantly lower complications and over 60% lower total low back pain-related costs. These studies have furthermore identified early PT as being associated with reduced rates of patient visits,

[15] Deyo, Richard, Sohail Mirza, Judith Turner, and Brook Martin. "Overtreating Chronic Back Pain: Time to Back Off?" J Am Board Fam Med 22(1) (2009): 62-68. doi:10.3122/jabfm.2009.01.080102.

surgeries, imaging, injections, and opioid prescriptions for low back pain,[16,17] saving over $2,000 in expenditures per patient. Another study by Kazis et al found that early PT and other conservative chiropractic services resulted in lower probabilities of short-term and long-term opioid use among the patient cohort, which had been found to be both harmful to patient health and a substantial driver of low back pain costs.[18] Based on these findings, there's a substantial opportunity to improve both the cost efficiency and quality of care for low back pain patients by incentivizing early PT.

Extensive literature suggests that some clinicians routinely and unnecessarily conduct diagnostic tests for low back pain (e.g., magnetic resonance imaging [MRI]) in the absence of symptoms suggesting serious low back pain problems. Numerous studies have demonstrated that in these cases, MRIs and other diagnostic tests are both highly costly and ineffective at leading to improved patient outcomes. In one set of randomized controlled trials, Kochen et al (2009) found no significant differences in primary outcomes, quality of life, mental health, or patient-reported satisfaction between cases with and without immediate lumbar imaging (e.g., radiography, MRI, CT). Meanwhile evidence shows that these tests are substantial drivers of low back pain-related costs, with expenditures in one study an average of $13,816 higher for cases with an MRI compared with the control.[19] Despite this evidence of minimal improvements and exorbitant costs, a study from Weiner et al (2006)[20] found that 41% of the sample population received MRIs on the same day as their low back pain diagnosis, despite the absence of "red flags" for serious underlying conditions.

Similar issues of waste exist with other forms of low back pain treatment. In one study, 41% of patients received early opioid prescriptions, a median of 25 days after diagnosis, resulting in a mean cost of $7,211 for these patients, more than three times the mean cost for the control group. Similar waste may exist for other expensive surgical treatments. Kim et al found that among a study population of patients with low back pain, surgery was performed for only about 1% of patients, yet the costs of care in these cases accounted for nearly 30% of total patient costs during the study period.[21] A study using the Patterns of Pain triage classification system to appropriately direct patients experiencing low back pain reduced the average cost of care for each episode of low back pain to $1,453 compared to $2,334 for patients not directed to care using the triage system.[22]

Wasteful low back pain treatments and tests present a substantial opportunity to improve the value of care for patients with low back pain. Multiple clinical guidelines currently suggest that clinicians should avoid conducting expensive tests (e.g., MRIs) within the initial few weeks of treatment, especially in the absence of any "red flags" for serious underlying conditions.

---

[16] Fritz et al. "Physical Therapy for Acute Low Back Pain: Associations with Subsequent Healthcare Costs" in Spine 33(16) (2008): 1800-1805. doi: 10.1097/BRS.0b013e31817bd853

[17] Fritz et al. "Primary Care Referral of Patients with Low Back Pain to Physical Therapy: Impact on Future Health Care Utilization and Costs" in Spine 37(25) (2012): 2114-2121. doi: 10.1097/BRS.0b013e31825d32f5

[18] Kazis LE, Ameli O, Rothendler J, Garrity B, Cabral H, McDonough C, Carey K, Stein M, Sanghavi D, Elton D, Fritz J, Saper R. Observational retrospective study of the association of initial healthcare provider for new-onset low back pain with early and long-term opioid use. BMJ Open. 2019 Sep 20;9(9):e028633. doi: 10.1136/bmjopen-2018-028633. Erratum in: BMJ Open. 2020 Jan 10;10(1):e028633corr1. PMID: 31542740; PMCID: PMC6756340.

[19] Weber et al., "The diagnostic utility of MRI in spondyloarthritis: An international multicentre evaluation of 187 subjects (The MORPHO study)" in Arthritis and Rheumatism Vol 62 Issue 10 (2010): https://doi.org/10.5167/uzh-34330

[20] See footnote 9.

[21] Kim et al. "Expenditures and Health Care Utilization Among Adults With Newly Diagnosed Low Back and Lower Extremity Pain." JAMA Netw Open 2(5) (2019): e193676. Doi:10.1001/jamanetworkopen.2019.3676

[22] Hall H, Prostko ER, Haring K, Fischer M, Cheng BC. A successful, cost-effective low back pain triage system: a pilot study. N Am Spine Soc J. 2021 Feb 1;5:100051. doi: 10.1016/j.xnsj.2021.100051. PMID: 35141617; PMCID: PMC8819953.

However, incentives are needed to balance potential risks incurred by clinicians from failing to perform these tests. Thus, cost measurement of these inefficient and expensive low back pain services has the opportunity to encourage clinicians to follow this clinical guidance, and as a result, promote high quality and efficient care in this clinical area.

2.2.2 Performance Scores

Table 1 shows the distribution of the measure score for clinician groups identified by a Tax Identification Number (TIN) and individual clinicians identified by a combination of a Tax Identification Number and National Provider Identifier (TIN-NPI).

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's room for improvement in the costs of care for a low back pain episode.

Table 1. Distribution of the Measure Score

Metric	TIN	TIN-NPI
Count	37,307	46,326
Mean Score	$1,710	$1,712
Score Standard Deviation	$517	$518
Minimum Score	$376	$395
Maximum Score	$7,489	$10,179
Score Interquartile Range (IQR)	$594	$617
Score Percentile		
10th	$1,153	$1,146
20th	$1,306	$1,300
30th	$1,427	$1,420
40th	$1,532	$1,531
50th	$1,640	$1,640
60th	$1,753	$1,758
70th	$1,885	$1,896
80th	$2,058	$2,074
90th	$2,330	$2,349

2.2.3 Disparities

Data on how the measure, as specified, addresses disparities is described in Sections 3.1.7 and 3.5.5.

3.0 Scientific Acceptability

3.1 Data Sample Description

Testing is based on the full population of measured entities with a minimum of 20 episodes and patients meeting inclusion and exclusion criteria for the measure, not based on a sample.

3.1.1 Type of Data Used for Testing

Medicare administrative claims, Long-Term Minimum Data Set (MDS), Medicare Enrollment Database (EDB), and Common Medicare Environment (CME).

3.1.2 Specific Dataset Used for Testing

The Low Back Pain measure uses Medicare Part A, B, and D claims data maintained by CMS. Claims data are used to build episodes of care, calculate episode costs, and construct risk adjustors. Episode costs are payment standardized and risk adjusted to ensure accurate comparison of cost across clinicians. Payment standardization adjusts the allowed amount for a Medicare service to limit observed differences in costs to those that may result from health care delivery choices. Data from the EDB are used to determine beneficiary-level exclusions and secondary risk adjustors, specifically Medicare Parts A, B, and C enrollment, primary payer, disability status, end-stage renal disease (ESRD), patient birth dates, and patient death dates. The risk adjustment model also accounts for expected differences in payment for services provided to patients in long-term care based on data from the MDS. Specifically, the MDS is used to create the long-term care indicator variable in risk adjustment.

3.1.3 Dates of the Data Used in Testing

Low Back Pain episodes ending from January 1, 2019, through December 31, 2019.

3.1.4 Levels of Analysis Tested

The measure was tested at group/practice (TIN) and individual clinician (TIN-NPI) levels.

3.1.5 Entities Included in the Testing and Analysis

Table 2 shows the individual clinician (identified by combination of TIN and NPI) and clinician group/practice (identified by TIN) included in the testing of the Low Back Pain measure.

Table 2. Measured Entities Characteristics with 20 Cases or More

Metric	TIN		TIN-NPI	
No Data	Count	%	Count	%
Count	37,307	100.00%	46,326	100.00%
Number of Episodes Attributed	-	-	-	-
20-39 Episodes	15,484	41.50%	22,456	48.47%
40-59 Episodes	7,525	20.17%	9,426	20.35%
60-79 Episodes	4,203	11.27%	5,099	11.01%
80-99 Episodes	2,482	6.65%	3,105	6.70%
100-199 Episodes	4,534	12.15%	5,051	10.90%
200-299 Episodes	1,201	3.22%	881	1.90%
300+ Episodes	1,878	5.03%	308	0.66%
Census Region	-	-	-	-
Northeast	7,093	19.01%	8,341	18.01%
Midwest	9,221	24.72%	11,278	24.34%

Metric	TIN		TIN-NPI	
No Data	Count	%	Count	%
South	12,789	34.28%	17,323	37.39%
West	8,101	21.71%	9,299	20.07%
Unknown	103	0.28%	85	0.18%

3.1.6 Patient Cohort Included in the Testing and Analysis

Table 3 shows the patient population for the Low Back Pain measure testing. It consists of Medicare beneficiaries enrolled in Medicare Parts A and B who receive care for low back pain that triggers a Low Back Pain episode.

Table 3. Beneficiary Demographics

Metric	Value
Count	3,155,095
Mean Age	72.30
Female %	60.8%
Part D %	75.45%

3.1.7 Social Risk Factors Included in Analysis

The analysis on social risk factors (SRFs) focused on examining the impact of Dual Medicare and Medicaid enrollment status on the measure. Table 4 outlines variables that may indicate SRFs and their advantages and disadvantages as indicators of individual-level SRFs. On balance, the analysis used dual Medicare and Medicaid enrollment status as the proxy of SRFs due to their broad availability in claims data, accurate measurement at the individual level, and wide acceptance of being a powerful indicator of health outcomes.[23]

Table 4. Social Risk Factors Available for Analysis

Variable	Advantages	Disadvantages	Used in Testing
Dual Medicare and Medicaid enrollment status	• Available for all beneficiaries • Most powerful predictor of poor outcomes[24]	• Variation in Medicaid eligibility across states	Yes

[23] 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

[24] See footnote 4.

Variable	Advantages	Disadvantages	Used in Testing
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[25] • Only 5 categories available, which may lack granularity to fully capture disparities[26,27]	No
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[28]	No
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 (AHRQ) index, deprivation index)	• Only a proxy measure, not always accurate at individual-level	No

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, group/practice (TIN) and individual clinician (TIN-NPI) levels.

3.2.2 Method of Reliability Testing

Data Element Reliability

The Low Back Pain 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.

- First, CMS routinely conducts data analyses 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.

[25] See footnote 4.

[26] 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).

[27] 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).

[28] 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

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.

- Second, 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. CMS continues to perform corrective actions and give providers additional education to ensure accurate billing.
- Lastly, to ensure claims completeness and inclusion of any corrections, the measure was developed and tested using data with a three-month claim run-out from the end of the measurement period.

Clinician-level Reliability

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

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

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

Where:

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

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

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

3.2.3 Statistical Results from Reliability Testing

Data Element Reliability

Between 2005 and 2019, CERT estimates that proper payment, which includes payments that met Medicare coverage, coding, and billing rules, ranged from 87.3% to 96.4% of total payments each year.[29] The fiscal year 2020 Medicare fee-for-service program proper payment rate was 93.7%.[30]

[29]Comprehensive Error Rate Testing (CERT) Program. "Appendices Medicare Fee-for-Service 2020 Improper Payments Report". Table A6. https://www.cms.gov/files/document/2020-medicare-fee-service-supplemental-improper-payment-data.pdf-1.
[30]Ibid.

Clinician-level Reliability

Table 5. Reliability at the Accountability Entity Level

Reporting Level	Entities Meeting Case Minimum (20 Episodes)	Mean Reliability	Median Reliability	% Above 0.4	% Above 0.7
TIN	37,307	0.752	0.784	96.27%	67.75%
TIN-NPI	46,326	0.729	0.761	95.66%	63.20%

3.2.4 Interpretation

The results of the data element testing show very high reliability of the critical data elements used by the measure for a 20-episode case minimum/volume threshold. At the accountability entity level, the measure is highly reliable for both the TIN and TIN-NPI reporting levels, at a mean reliability of 0.752 and 0.729 respectively. For reference, CMS generally considers 0.4 as the threshold indicating 'moderate' reliability and 0.7 as high reliability.[31] Additionally, the vast majority of TINs and TIN-NPI meet or exceed the moderate reliability threshold of 0.4 and most are above the high reliability threshold of 0.7.

3.3 Validity Testing

3.3.1 Level of Validity Testing

The validity of the measure was tested using face validity and empirical validity at group/practice (TIN) and individual clinician (TIN-NPI) levels.

3.3.2 Method of Validity Testing

Face Validity

The Low Back Pain measure was developed through a structured, iterative process for gathering detailed input from recognized clinician experts on the measure. Experts in this clinical area evaluated specifications to ensure that each aspect of the measure (e.g., assigned services) was intentionally capturing only the costs of care within the reasonable influence of the attributed clinician for a defined patient population (i.e., the ability of the measure score to differentiate good from poor performance).

In developing this measure, Acumen incorporated input from:

(i) a Low Back Pain Clinician Expert Workgroup;
(ii) a Technical Expert Panel (TEP); and
(iii) the Person and Family Partners.

This process is detailed in the Episode-Based Cost Measures Development Process document posted on the MACRA Feedback Page.[32]

[31] 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.
[32] CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.

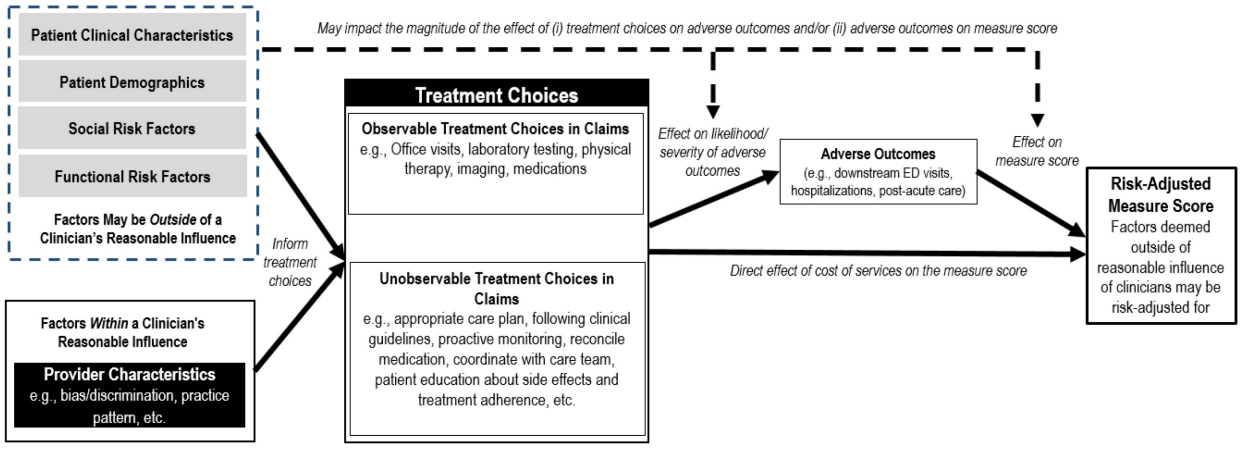
One of the key roles of the measure-specific Clinician Expert Workgroup is to develop service assignment rules for the cost measure. These service assignment rules are intended to ensure clinicians are evaluated on services and costs that are clinically related to the attributed clinician's role in treating and managing the condition, thus limiting cost variation unrelated to clinician care this measure. Therefore, assigned services are services that the Clinical Expert Workgroup believed an attributed clinician can influence their occurrence, frequency, or intensity.

Prior to submitting the measure for the Measure Under Consideration list, members of the Clinician Expert Workgroup were asked to consider the measure as specified and rate the degree to which the actions outlined in the logic model is within the reasonable influence of an attributed clinician, and by extension, can affect patient health outcomes and downstream costs.

Empirical Validity Testing

We evaluated the empirical validity of the Low Back Pain measure by estimating the effect of relevant treatment choices on the measure score using multiple regression, based on the conceptual model outlined in Figure 2. For more information on the conceptual model, please see Section 3.5.3.

Figure 2: Conceptual Model of the Relationship between Treatment Choices and the Measure Score



The cost measure is designed to reflect cost directly related to treatment choices, as well as 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 outcome, in turn, contributes to the total cost 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 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.

3.3.3 Statistical Results from Validity Testing

Face Validity

Figures 3 to 7 show the responses of the Clinical Expert Workgroup Members, when asked to consider the measure as specified and rate the degree to which the actions by an attributed clinician outlined in the logic model are within their reasonable influence and can affect patient health outcomes and downstream costs.

Figure 3. Reponses of Clinical Expert Workgroup Members when Asked to Rate the Degree of Influence of Attributed Clinicians over Actions Outline in the Logic Model

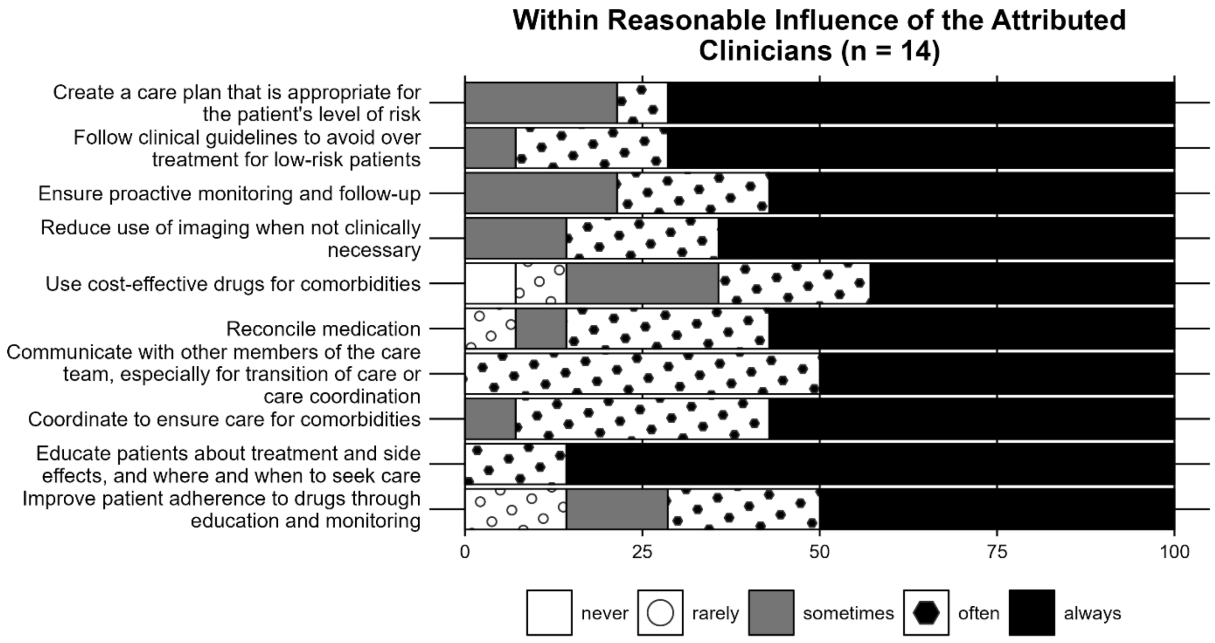


Figure 4. Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Impact on Risk of High-Cost Events for Actions Outline in the Logic Model

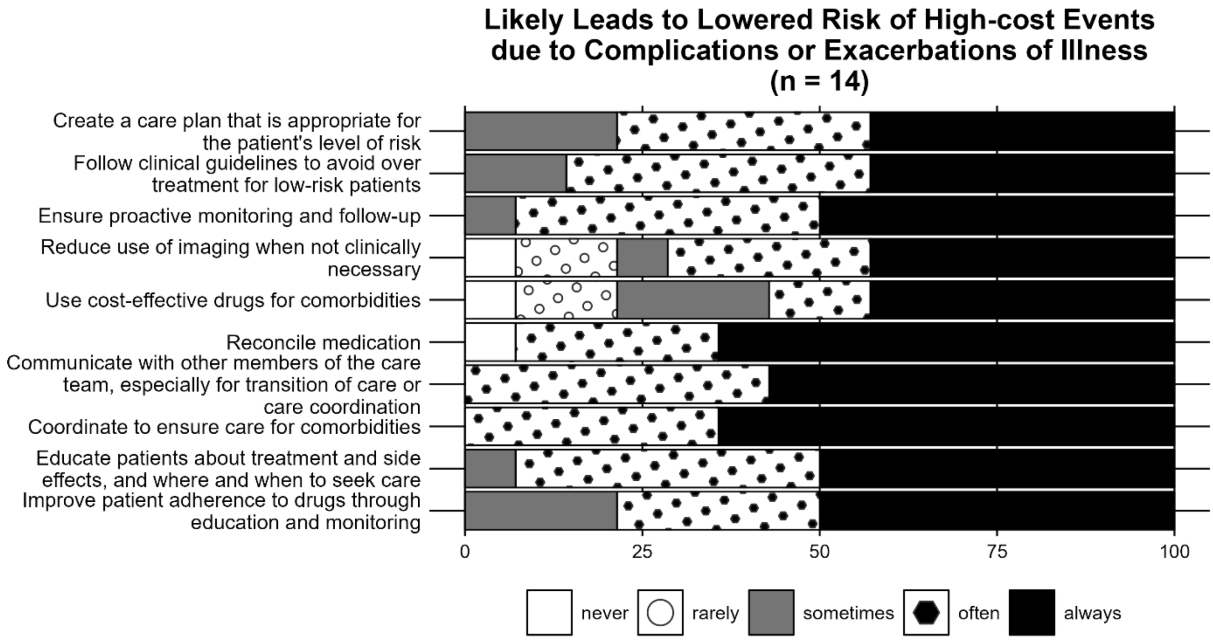


Figure 5. Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Improving Patient Treatment and Quality of Life for Actions Outline in the Logic Model

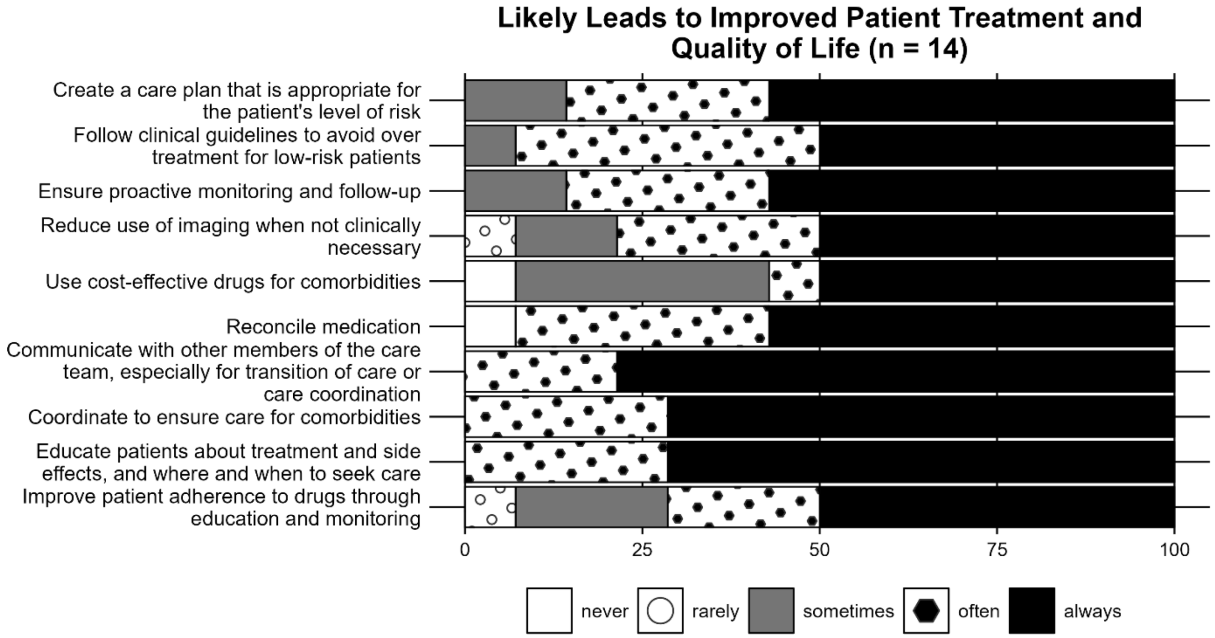


Figure 6. Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Improving Monitoring and Care Coordination for Actions Outline in the Logic Model

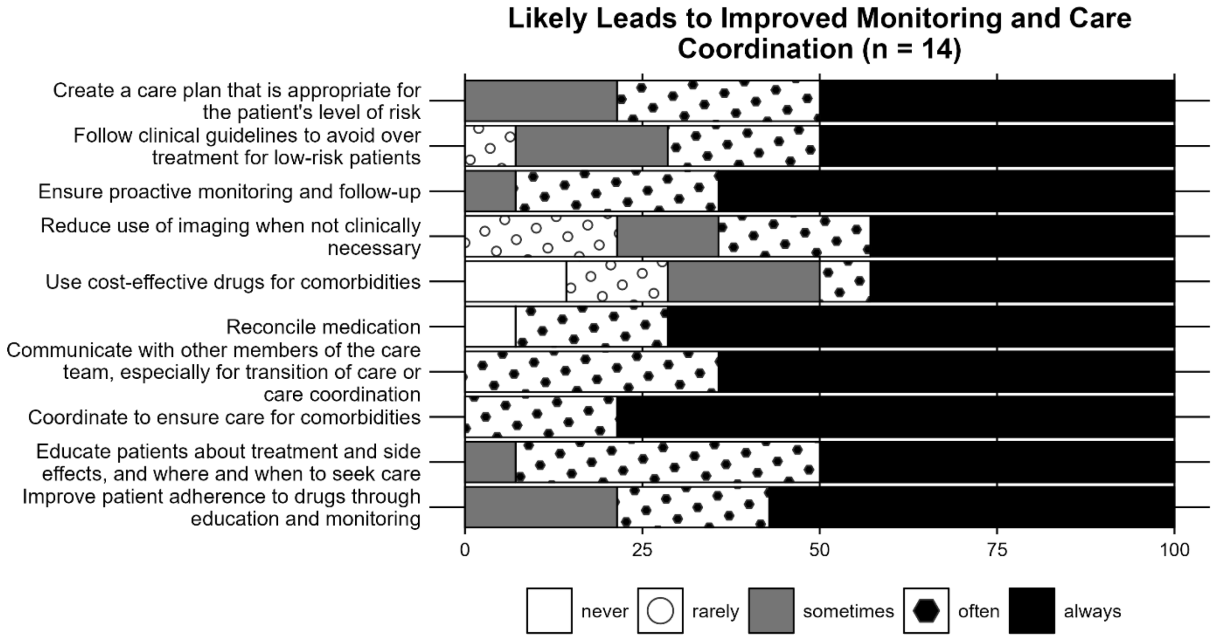
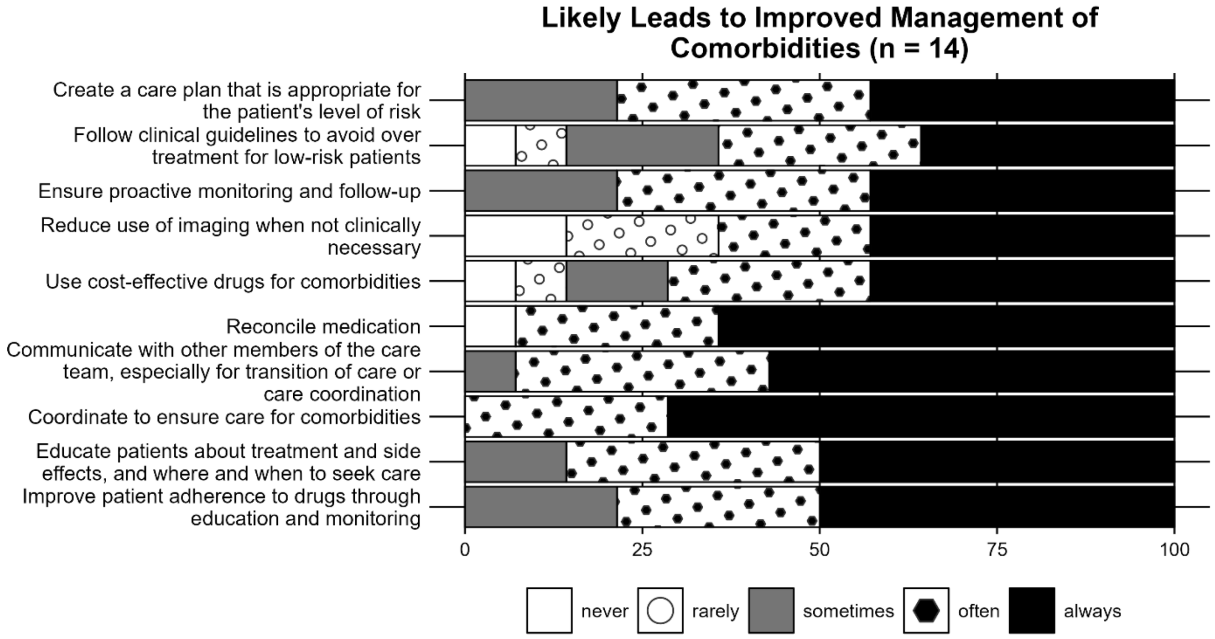


Figure 7. Reponses of Clinical Expert Workgroup Members when Asked to Rate the Likelihood of Improving Management of Comorbidities for Actions Outline in the Logic Model



Empirical Validity Testing

Table 6 shows two regression models for each reporting level. Model 1 shows the effect on the clinicians' mean observed cost to expected cost ratio (O/E) for each additional one thousand dollars of a service category that is assigned to an episode, on average, while holding the remaining categories of cost constant. Model 2 shows the effect on the mean cost of adverse events for each additional one thousand dollars of a cost category that is assigned to an episode, on average, while holding the remaining categories of services constant.

Table 6: Estimated Effect of Treatment Choices

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	0.43 [0.41, 0.44] (p < 0.01)	-	0.41 [0.40, 0.42] (p < 0.01)	-
Outpatient Evaluation & Management Services	-0.06 [-0.09, -0.04] (p < 0.01)	0.96 [0.94, 0.98] (p <0.01)	-0.04 [-0.06, -0.01] (p < 0.01)	0.64 [0.63, 0.65] (p < 0.01)
Ambulatory/Minor Procedures	0.25 [0.24, 0.26] (p < 0.01)	-0.04 [-0.05, -0.03] (p < 0.01)	0.25 [0.24, 0.25] (p < 0.01)	-0.01 [-0.01, 0.00] (p < 0.01)
Outpatient Physical, Occupational, or Speech and Language Pathology Therapy	0.56 [0.55, 0.57] (p < 0.01)	-0.08 [-0.09, -0.08] (p < 0.01)	0.64 [0.63, 0.65] (p < 0.01)	-0.09 [-0.09, -0.08] (p < 0.001)
Laboratory, Pathology, and Other Tests	0.41 [0.36, 0.46] (p < 0.01)	-0.31 [-0.36, -0.27] (p < 0.01)	0.43 [0.39, 0.47] (p < 0.01)	-0.07 [-0.10, -0.05] (p < 0.01)
Imaging Services	0.51 [0.46, 0.57] (p < 0.01)	-0.44 [-0.49, -0.40] (p <0.01)	0.45 [0.41, 0.49] (p < 0.01)	-0.21 [-0.24, -0.19] (p < 0.01)
Durable Medical Equipment and Supplies	0.76 [0.71, 0.81] (p < 0.01)	0.52 [0.48, 0.57] (p < 0.01)	0.56 [0.51, 0.06] (p < 0.01)	0.30 [0.27, 0.33] (p < 0.01)
Anesthesia Services	-0.53 [-0.66, -0.41] (p < 0.01)	0.13 [0.02, 0.23] (p = 0.02)	-0.35 [-0.44, -0.27] (p < 0.01)	0.44 [0.37, 0.50] (p < 0.01)
Chemotherapy and Other Part B-Covered Drugs	0.19 [0.18, 0.20] (p < 0.01)	-0.03 [-0.04, -0.02] (p < 0.01)	0.19 [0.18, 0.20] (p < 0.01)	0.00 [-0.01, 0.01] (p = 0.47)
Part-D Drugs	0.21 [0.18, 0.24] (p < 0.01)	-0.16 [-0.19, -0.13] (p < 0.01)	0.19 [0.17, 0.22] (p < 0.01)	-0.04 [-0.06, -0.02] (p < 0.01)

3.3.4 Interpretation

Face Validity

Overall, there's strong consensus among workgroup members that all of the actions outlined in the logic model are often or always within a reasonable influence of the attributed clinician (Figure 3) and are likely to lead to lowered risk of downstream high-cost events and improved treatment and patient outcomes if performed by the attributed clinician, with every action receiving above 50% of responses that rated often or always (Figures 4 and 5). Similarly, there's strong consensus among clinician experts that all of the actions outlined in the logic model can lead to improved monitoring and care coordination, as well as improved management of comorbidities with every action receiving above 50% of responses that rated often or always as shown in (Figures 6 and 7).

Empirical Validity

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 6). Therefore, there's evidence that the measure is capturing what it purports to measure.

The results are also consistent with performance gaps identified from the literature review in Section 2.2.1, such as overuse in imaging or physical therapy. Model 1 shows that having more adverse events is associated with worse scores, which includes hospitalizations, emergency department (ED) visits, and post-acute care.

Model 1 also shows that increasing cost of physical therapy, laboratory testing, imaging and drugs are associated with worse score. However, these categories of services are also associated with decreasing cost of adverse events in model 2, which suggest that the cost of these services can also indirectly influence the measure score through the cost of adverse events. For durable medical equipment cost, it appears to be associated with both worse score and increasing cost of adverse events, which suggests that the cost of durable medical equipment directly influence the measure score and the higher usage may be linked to risk of adverse events. Lastly, the cost of anesthesia services is shown to be associated with better score and increasing cost of adverse events, which is also reflective of higher service intensity that is potentially linked to risk of adverse events.

3.4 Exclusions Analysis

3.4.1 Method of Testing Exclusions

Exclusions are used in the Low Back Pain measure to ensure a comparable patient population within the scope of the measure's focus on patients who receive care for low back pain and that episodes provide meaningful information to attributed clinicians. Exclusions are also used as part of data processing so that sufficient data are available to accurately determine episode spending and calculate risk adjustment for each episode.

For the exclusions analysis discussed in this section, we focused on exclusion criteria intended to ensure a comparable patient population.

Standard exclusions are done to ensure data completeness. Episodes meeting the criteria listed below are excluded because they may not accurately reflect a clinician's performance in managing and treating patients with low back pain.

- The patient has a primary payer other than Medicare for any time overlapping the episode window or 120-day lookback period prior to the episode window
- The patient was not enrolled in Medicare Parts A and B for the entirety of the 120-day lookback period plus episode window, or was enrolled in Part C for any part of the 120-day lookback period plus episode window

- The patient isn't found in the Medicare Enrollment Database (EDB)
- The patient has an episode window shorter than 120 days
- The patient's death date occurred before the episode end date
- The patient resided outside the United States or its territories during the episode

This analysis also included exclusion criteria specific to the Low Back Pain episode-based cost measure. The analysis excluded episodes containing the following conditions or procedures that may have different care pathways or characteristics making them incomparable to the rest of the episodes.

- Cauda equina syndrome
- Myelopathy
- Osteoporotic Compression Fracture
- Spinal Infection
- Spinal Neoplasm
- Spine surgery performed during episode window (Post Trigger)
- Trauma

Given the rationales for these exclusions, we would expect these excluded episodes to have a different profile than the included episodes, such as a higher mean cost, or a different distribution of costs (e.g., a long tail of high-cost episodes). For each exclusion, we examined the number of episodes and beneficiaries affected, as well as the distributions of observed cost. We then compared the cost characteristics of the excluded episodes to those of episodes included in measure calculation to assess the distinctness between the 2 patient cohorts. A full list of the exclusions used for the Low Back Pain measure is provided in the Measure Codes List available on the MACRA Feedback Page.[33]

3.4.2 Statistical Results from Testing Exclusions

Table 7 below presents 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.

[33]CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.

Table 7. Cost Statistics for Measure Exclusions

Exclusion	Episodes		Observed Cost					
			Mean	Percentile				
	#	%		10th	25th	50th	75th	90th
All Episodes Meeting Triggering Logic	4,432,188	100.00%	$2,192	$141	$335	$853	$1,861	$4,115
Episode Length Less Than 120 Days	25,580	0.58%	$5,697	$439	$882	$2,031	$4,583	$9,653
Beneficiary Death in Episode	61,523	1.39%	$4,226	$331	$673	$1,583	$3,608	$7,666
Outlier	82,240	1.86%	$4,851	$183	$355	$4,987	$5,985	$9,685
Cauda equina syndrome	3,986	0.09%	$6,000	$449	$949	$2,130	$5,440	$13,878
Osteoporotic Compression Fracture	18,642	0.42%	$4,744	$346	$839	$2,037	$4,961	$10,291
Spinal Infection	12,733	0.29%	$5,453	$339	$813	$1,838	$4,484	$10,984
Myelopathy	99,165	2.24%	$3,554	$211	$539	$1,251	$2,675	$6,955
Spinal Neoplasm	4,779	0.11%	$5,433	$409	$883	$1,959	$4,786	$11,064
Spine Surgery Performed within 60 Days of the Trigger)	80,316	1.81%	$19,329	$1,212	$6,425	$11,345	$29,538	$44,350
Trauma	69,612	1.57%	$4,544	$346	$796	$1,823	$4,479	$9,593
TIN does not Meet Case Minimum	422,491	9.53%	$2,041	$140	$295	$732	$1,753	$4,047
No Attributed NPI	260,583	5.88%	$3,244	$477	$816	$1,460	$2,737	$5,985
TIN-NPI does not Meet Case Minimum	1,248,091	28.16%	$2,498	$193	$402	$944	$2,033	$4,853
Reportable Episodes (if all clinicians reported as TIN at the Testing Volume Threshold)	3,652,705	82.41%	$1,746	$134	$324	$817	$1,713	$3,286
Reportable Episodes (if all clinicians reported as TIN-NPI at the Testing Volume Threshold)	2,693,187	60.76%	$1,594	$112	$269	$718	$1,584	$3,030

3.4.3 Interpretation

Table 7 displays descriptive statistics of all episodes meeting the measure's triggering logic, measure-specific excluded episodes, and the final reportable episodes at the group- and individual level. The statistical results show that the distribution of observed costs for all episodes is higher than that of reportable episodes, with a mean of $2,192 for all episodes and $1,746 and $1,594 for reportable levels at the TIN and TIN-NPI level, respectively. All exclusion criteria have high mean observed episode cost than all episodes meeting triggering logic, which supports the exclusion of these episodes to ensure a comparable and clinically coherent patient cohort that will yield a clinically coherent measure and meaningful information to attributed clinicians.

3.5 Risk Adjustment or Stratification

3.5.1 Method of Controlling for Differences

Differences in case mix are controlled for using a statistical risk model with 162 risk factors and stratification by 8 risk categories.

The risk adjustment model for the Low Back Pain measure adjusts for comorbidities based on the CMS Hierarchical Condition Category (HCC) model, count of HCCs, end-stage renal disease (ESRD) status, disability status, number and types of clinician specialties from which the patient has received care, recent use of institutional long-term care, and age.

The model also includes measure-specific factors:

- Medical back problems hospitalization
- Cognitive status/dementia
- Depression
- Fibromyalgia
- Frailty
- Osteoarthritis
- History of opioid use
- Osteoporosis
- Spine surgery performed during lookback period (1 year)
- Smoking
- Scoliosis and Other Spinal Deformities
- Spondylolysis

A separate linear regression is run for each sub-group and Medicare Part D enrollment status combination to ensure fair comparison:

- Surgical episode with history of complex low back pain with Part D enrollment
- Surgical episode without history of complex low back pain with Part D enrollment
- Non-surgical episode with history of complex low back pain with Part D enrollment
- Non-surgical episode without history of complex low back pain with Part D enrollment
- Surgical episode with history of complex low back pain without Part D enrollment
- Surgical episode without history of complex low back pain without Part D enrollment
- Non-surgical episode with history of complex low back pain without Part D enrollment
- Non-surgical episode without history of complex low back pain without Part D enrollment

The episode's scaled (i.e., annualized) observed costs are winsorized at the 98th percentile prior to the regression for each model to handle extreme observations. Full details of the risk adjustment model are in the Measure Codes List File available on the MACRA Feedback page.[34]

3.5.2 Conceptual, Clinical, and Statistical Methods

We selected the CMS-HCC model based on previous studies evaluating its appropriateness for use in risk adjusting Medicare claims data. This model was developed specifically for use in the Medicare population, meaning that it accounts for conditions found in the Medicare population. In addition, the CMS-HCC model is routinely updated for changes in coding practices (e.g., the transition from ICD-9 to ICD-10 codes). Because the CMS-HCC model has already been

[34]CMS, MACRA Feedback Page, https://www.cms.gov/Medicare/Quality-Payment-Program/Quality-Payment-Program/Give-Feedback.

extensively tested, we focus our testing on the adaptation of the CMS-HCC model to the Low Back Pain measure's patient population.

The workgroup provided input on measure-specific risk adjustors after reviewing empirical analyses on subpopulations of interest to assess whether and if so, how, particular factors should be accounted for in the model. These could include patient characteristics, factors outside of the reasonable influence of the clinician, or any other factors that would help prevent unintended consequences. These additional risk adjustors are listed in the section above.

As previously noted, the risk adjustment model is run on episodes stratified into episode sub-groups, which may qualify as "ordering" of risk factors. Episode sub-groups were also determined based on the workgroup's input, with the goal of ensuring clinical comparability among episodes so that the cost measure fairly compares clinicians with similar patient case-mix.

3.5.3 Conceptual Model of Impact of Social Risks

Figure 2 in Section 3.3.2 shows the conceptual model that outlines how SRFs can influence the measure score, which is informed by both published external research and our own data analysis.[35,36,37,38,39] The conceptual model outlines risk factors that are either known by the literature or informed by the Clinical Expert Workgroup to be within or outside of influence of the attributed clinician. Risk factors, including SRFs, can both influence the treatment choices and impact the size of the effect of treatment choices on mitigating 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, 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 HCCs. 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 based on those discussions and consensus among the clinical experts. Additionally, during our testing phases, we also follow a structured and systematic approach to decide whether SRFs should be adjusted for, which is further described in Section 3.5.5.

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., Accountable Care Organizations, previous physician Quality and Resource Use Report programs, and other administrative claims-based measures such as the Knee Arthroplasty episode-based cost

---

[35] See footnote 15.

[36] 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.

[37] 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

[38] 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/.

[39] 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/social-risk-factors-and-medicares-value-based-purchasing-programs

measure, Total Per Capita Cost (TPCC) cost measure, Medicare Spending Per Beneficiary (MSPB)-PAC cost measure and MSPB Hospital cost measure). Recalling that the risk model relies on the existing CMS-HCC model, testing results for factors included in the CMS-HCC V22 2016 model can be found in the Evaluation of the CMS-HCC Risk-Adjustment Model report[40] and the Report to Congress: Risk Adjustment in Medicare Advantage[41]. For measure-specific factors not included in the CMS-HCC model, we sought expert clinician input through the workgroup, which provided recommendations on additional risk adjustors and measure sub-groups.

3.5.5 Analyses and Interpretation in Selection of Social Risk Factors

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 clinician's fixed effects,
(iii) whether 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 does not risk adjust for SRFs.

Overall, the results suggest that it isn't appropriate to risk adjust for social risk factors in this measure. There's a statistically significant association between the patient's dual status and episode cost in some subgroups (Table 8). However, this association isn't stable and no longer statistically significant in many subgroups after adding variables to account for provider-level factors, which suggests that the patient-level factors are less influential than provider-level factors. This is also supported by the fact that the performance degradation with increasing shares of dual patients is more accelerated for dual episodes than non-dual episodes (Table 9). Many providers are able to mitigate the effect of social risk by performing equally well or significantly better on their dual episodes (Table 10). Lastly, risk adjusting for dual status does not appear to substantially change the performance ranking for many providers (Table 11).

Table 8: Coefficient of Patient-level Dual Status under Different Models

[40]Pope, Gregory C., John Kautter, et al., "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.
[41]CMS, "Report to Congress: Risk Adjustment in Medicare Advantage," https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/RTC-Dec2018.pdf.

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	Non-surgical episode with history of complex low back pain with Part D Coverage	33.78%	$85 (p <0.01)	$47 (p < 0.01)	$63 (p < 0.01)
TIN	Non-surgical episode with history of complex low back pain without Part D Coverage	10.36%	$190 (p < 0.01)	$177 (p < 0.01)	$182 (p < 0.01)
TIN	Non-surgical episode without history of complex low back pain with Part D Coverage	39.56%	$72 (p < 0.01)	$44 (p < 0.01)	$51 (p < 0.01)
TIN	Non-surgical episode without history of complex low back pain without Part D Coverage	13.33%	$76 (p = 0.001)	$71 (p = 0.003)	$71 (p = 0.005)
TIN	Surgical episode with history of complex low back pain with Part D coverage	1.61%	$538 (p = 0.01)	$820 (p < 0.01)	$493 (p = 0.049)
TIN	Surgical episode with history of complex low back pain without Part D coverage	0.53%	$53 (p = 0.98)	$231 (p = 0.91)	-$1,901 (p = 0.502)
TIN	Surgical episode without history of complex low back pain with Part D coverage	0.62%	$852 (p = 0.01)	$1,234 (p < 0.01)	$1,371 (p < 0.01)
TIN	Surgical episode without history of complex low back pain without Part D coverage	0.22%	-$2,275 (p = 0.46)	-$2,232 (p = 0.47)	-$1,390 (p = 0.732)
TIN-NPI	Non-surgical episode with history of complex low back pain with Part D Coverage	33.79%	$92 (p < 0.01)	$65 (p < 0.01)	$80 (p < 0.01)
TIN-NPI	Non-surgical episode with history of complex low back pain without Part D Coverage	10.35%	$203 (p < 0.01)	$204 (p < 0.01)	$227 (p < 0.01)
TIN-NPI	Non-surgical episode without history of complex low back pain with Part D Coverage	39.72%	$79 (p < 0.01)	$63 (p < 0.01)	$68 (p < 0.01)
TIN-NPI	Non-surgical episode without history of complex low back pain without Part D Coverage	13.38%	$79 (p < 0.01)	$84 (p < 0.01)	$90 (p < 0.01)

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-NPI	Surgical episode with history of complex low back pain with Part D coverage	1.53%	$475 (p = 0.04)	$830 (p < 0.01)	$628 (p = 0.065)
TIN-NPI	Surgical episode with history of complex low back pain without Part D coverage	0.50%	-$327 (p = 0.89)	-$18 (p = 0.99)	-$1,976 (p = 0.678)
TIN-NPI	Surgical episode without history of complex low back pain with Part D coverage	0.54%	$1,072 (p < 0.01)	$1,368 (p < 0.01)	$2,290 (p < 0.01)
TIN-NPI	Surgical episode without history of complex low back pain without Part D coverage	0.19%	NA	NA	NA

Table 9: Mean Ratio of Observed Cost to Expected Cost (O/E) Stratified by Clinician's Dual Share and Patient's Dual Status

Dual Share	TIN			TIN-NPI		
	All Episode	Dual Episodes	Non-Dual Episodes	All Episodes	Dual Episodes	Non-Dual Episodes
All	0.99	1.03	0.99	1.01	1.04	1.00
0%	0.96	-	0.96	0.98	-	0.98
1-20%	0.99	1.02	0.98	1.00	1.04	1.00
21-40%	1.03	1.04	1.02	1.03	1.06	1.02
41-60%	1.01	1.02	1.00	1.00	1.01	0.98
61-80%	1.08	1.10	1.04	1.04	1.06	0.99
81-99%	1.16	1.16	1.11	1.16	1.16	1.11
100%	1.35	1.35	-	1.31	1.31	-

Table 10: Proportions of Clinicians Who Perform Significantly Worse Equally Well, or Significantly Better on Their Dual Episodes than Non-Dual Episodes

Reporting Level	Significantly Worse	Equally Well	Significantly Better
TIN	6.69%	91.76%	1.55%
TIN-NPI	6.55%	92.47%	0.98%

Table 11: Clinicians' Performance Shift after Adding a Dual Status Risk Adjustor

TIN or TIN-NPI	Proportion of Clinicians Affected at Various Levels of Performance Shift	
	Ranking Shift by 1% or more	Ranking Shift by 5% or more
TIN	40.72%	1.05%
TIN-NPI	41.87%	1.26%

3.5.6 Method for Statistical Model or Stratification Development

To analyze the validity of current risk adjustment model, we examined two criteria: discrimination and calibration.

1) 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. These results are provided in Section 3.5.7.
2) 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. A well-calibrated measure should have predictive ratios close to 1.0 across all deciles. These are discussed in Sections 3.5.8 and 3.5.9.

3.5.7 Statistical Risk Model Discrimination Statistics

The overall R-squared for the Low Back Pain cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.53. The adjusted R-squared is 0.53. More information on discrimination testing for the CMS-HCC model can be found at Pope et al. 2011.[42]

3.5.8 Statistical Risk Model Calibration Statistics

The predictive ratio is calculated using the formula of average expected cost / average observed cost for all episodes in each decile.

3.5.9 Statistical Risk Model Calibration – Risk Decile

Analysis of predictive ratios by risk decile for the measure shows minor variation among risk deciles, as predictive ratios range from 0.95 to 1.03 across all risk deciles (with an overall average of 1.00). The average predictive ratios for deciles 1 through 3 are slightly further from 1 than the rest, and there's a jump from underestimating the cost in the 1st decile to overestimating the cost in the 2nd and 3rd deciles.

[42]Pope, Gregory C., John Kautter, et al., "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

Table 12: Predictive Ratio by Decile of Predicted Episode Cost

Decile	Average Predictive Ratio
Decile 1	0.95
Decile 2	1.03
Decile 3	1.03
Decile 4	0.98
Decile 5	0.98
Decile 6	0.99
Decile 7	1.00
Decile 8	1.01
Decile 9	1.01
Decile 10	1.00

3.5.10 Interpretation

The R-squared values for the model, which measure the percentage of variation in results predicted by the model, are higher than the values presented in similar analyses of risk adjustment models.[43] As noted in Section 3.5.6 and 3.5.7, these results should be interpreted alongside service assignment rules, which remove clinically unrelated services.

The remaining unexplained variance is due to variation in factors that aren't 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 isn't 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.

Table 12 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 0.95 and 1.03. Overall, the risk adjustment model does not over- or under-predict cost across the full range of resource use patterns in the population.

3.6 Identification of Meaningful Differences in Performance

3.6.1 Method

To identify meaningful differences in performance, this analysis first examines the distribution of the measure score to highlight the performance gap between the most and least efficient clinicians. Then, this analysis examines the rate of high-cost events that may occur during an episode of care to highlight the variation in frequency and cost of those events.

3.6.2 Statistical Results

Table 1 shows the distribution of the measure score at the TIN and TIN-NPI levels. The rate of a clinically related emergency department visit during an episode is 6.85% and the rate of a clinically related hospitalization during an episode is 2%.

[43] Ibid.

3.6.3 Interpretation

There's substantial variation observed in the measure score in both TIN and TIN-NPI levels, indicated by the interquartile ranges, standard deviations, and coefficients of variation. The magnitude of the observed variation is in the thousands of dollars, which indicates that there are opportunities to close the gaps between the most and least efficient clinicians.

There are also opportunities to reduce costs associated with high-cost events, such as clinically related emergency department visits and hospitalizations. Episodes with a clinically related emergency department visit cost Medicare approximately $346 million more than an average episode, and $92.2 million for episodes with a clinically related acute inpatient stay.

3.7 Missing Data Analysis and Minimizing Bias

3.7.1 Method

Since CMS uses Medicare claims data to calculate the Low Back Pain measure, Acumen expects a high degree of data completeness. To further ensure that we've complete and accurate data for each patient, Acumen excludes episodes where patient date of birth information (an input to the risk adjustment model) can't be found in the EDB, the patient does not appear in the EDB, or the patient death date occurs before the episode trigger date.

The Low Back Pain measure also excludes episodes where the patient is enrolled in Medicare Part C or has a primary payer other than Medicare in the 120-day lookback period and episode window. In such situations, Medicare Parts A and B claims data may not capture the complete clinical profile for the patient needed to capture the clinical risk of the patient in risk adjustment. Furthermore, Parts A and B claims data may not capture all Medicare resource use if some portion of the patient'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 categories of missing data which caused episodes to be excluded from the Low Back Pain measure. Frequency is presented in terms of the number of episodes excluded due to missing data, as well as the cost profile of episodes with missing data compared to episodes included the measure reporting.

As a note, the episode and clinician counts below reflect exclusion from the initial population of triggered episodes. After the missing data exclusions are applied, we then apply additional exclusions, as outlined in Section 3.4, to this overall patient cohort to narrow the population to only applicable episodes.

Table 13: Cost Statistics for Missing Data Category

Missing Data Categories	Episodes	Observed Cost					
		Mean	Percentile				
			10th	25th	50th	75th	90th
Beneficiary Resides Outside of U.S. or Territories	6,953	$1,822	$189	$345	$706	$1,461	$2,794

Missing Data Categories	Episodes	Observed Cost					
		Mean	Percentile				
			10th	25th	50th	75th	90th
Primary Payer Other than Medicare	742,255	$2,188	$152	$343	$824	$1,777	$3,753
No Continuous Enrollment in Medicare Parts A and B, and Any Enrollment in Part C	702,718	$1,894	$98	$239	$647	$1,518	$3,257
Reportable Episodes - Group Reporting	3,652,705	$1,746	$134	$324	$817	$1,713	$3,286
Reportable Episodes - Individual Reporting	2,693,187	$1,594	$112	$269	$718	$1,584	$3,030

3.7.3 Interpretation

The results show that the missing data episodes don't appear to be substantially different than all episodes in the initial population in terms of cost (Table 13). Therefore, the impact of removing these episodes on the overall measure should be minimal while ensuring that clinicians are fairly evaluated on episodes with complete data.

4.0 Feasibility

4.1 Data Elements Generated as Byproduct of Care Processes

The data elements used in this measure are pulled from Medicare claims. They can be based on information generated, collected and/or used by healthcare personnel during the provision of care (e.g., diagnoses), which are then translated into the appropriate coding system (e.g. ICD-10 diagnoses, MS-DRGs) for use in Medicare claims by either the original healthcare personnel or another individual.

4.2 Electronic Sources

All data elements are in defined fields in electronic claims.

4.3 Data Collection Strategy

4.3.1 Data Collection Strategy Difficulties

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

4.3.1.1 Data Collection

Acumen receives claims data directly from the Common Working File (CWF) maintained at the CMS Baltimore Data Center. Medicare claims are submitted by healthcare providers to a Medicare Administrative Contractor (MAC), and are subsequently added to the CWF. However, these claims may be denied or disputed by the MAC, leading to changes to historical CWF data. In rare circumstances, finalizing claims may take many months, or even years. As a result, it isn't practical to wait until all claims for a given month are finalized before calculating this measure. As such, there's 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 is used in a CMS program, calculation and reporting would be done in line with that program's reporting practices.

4.3.1.2 Missing Data

This measure requires complete beneficiary information, and a small number of episodes with missing data are excluded to ensure completeness of data and accurate comparability across episodes. For example, episodes where the beneficiary was not enrolled in Medicare Parts A and B for the 120 days prior to the episode start date aren't included in this measure. This enables the risk adjustment model to accurately adjust for the beneficiary's comorbidities using data from the previous 120 days of Medicare claims. Additionally, the risk adjustment model includes a categorical variable for beneficiary age bracket, so episodes for which the beneficiary's date of birth can't be located aren't included in this measure.

4.3.1.3 Sampling

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

5.0 Usability and Use

5.1 Use

5.1.1 Current and Planned Use

The Low Back Pain measure isn't currently in use, but is intended for use in a payment program and could eventually be publicly reported. The measure was specifically developed for potential use in the Cost performance category of MIPS to assess clinicians reporting as individuals or groups, under a contract with CMS.

For the measure to be used in MIPS, it must be reviewed by the Measure Application Partnership (MAP) and then undergo the notice-and-rulemaking process. Given these next steps, the earliest the measure could be in use in MIPS is CY 2024. If in use, CMS can then determine whether to publicly report the cost measure.

5.1.2 Feedback on the Measure by Those being Measured or Others

Throughout the Low Back Pain measure development, we used an iterative and extensive process to gather feedback on the measure and its results to ensure that the measure can be used appropriately in the MIPS program by clinicians and clinician groups who practice in this clinical area. This process also aims to make sure that the measure performance results can be understood by the population that is being measured to help support decision making. A couple of the main ways that we gathered feedback was through i) reoccurring Clinician Expert Workgroup meetings, where members discussed the clinical perspective, the patient perspective, and empirical data, in order to recommend measure specifications, and ii) the national field testing of the measure.

5.1.2.1 Technical Assistance Provided During Development or Implementation

Clinician Expert Workgroup Meetings

For each Clinician Expert Workgroup meeting, Acumen provided empirical data (e.g., analyses on potentially relevant services to group and potential sub-populations to sub-group, risk adjust, or exclude). These analyses were conducted using all administrative claims data for Medicare Parts A, B, and D. This data was shared with Workgroup members to help inform their feedback on the measure specifications throughout its development to ensure that the measure was appropriately assessing costs for the attributed clinicians.

Field Testing

Additionally, Acumen and CMS nationally field tested the draft Low Back Pain measure, along with 4 other episode-based cost measures, for a 10-week comment period (January 10 to March 25, 2022). We provided a Field Test Report with performance data to all clinician groups and clinicians who were attributed 20 or more episodes.[44] This testing sample was selected to balance coverage and reliability, since a key goal of field testing was to test the measures with as many clinicians and other interested members of the public as possible. A total of 49,949 TIN and 69,742 TIN-NPI reports were developed for this measure. During this time, feedback was gathered on the usability of the performance data and the appropriateness of the measure.

[44]The field test reports were available for download from the Quality Payment Program website: https://qpp.cms.gov/login.

5.1.2.2 Technical Assistance with Results

Clinician Expert Workgroup Meetings

Acumen provided data in advance of or during each of the Clinician Expert Workgroup Meetings: Workgroup meeting, Service Assignment and Refinement Meeting, Post-Field Test Refinement Meeting. During the meetings, Acumen would guide Workgroup members through these analyses, providing clinical and programmatic context when needed. Using this iterative process, the Workgroup members discussed the testing results in depth during each meeting and allowed the data to inform their recommendations for measure specifications. The goal was to ensure that the measure was appropriately assessing clinicians cost of care within their reasonable influence, without creating potential unintended consequences so that it could be usable in the MIPS program.

Field Testing

During the field testing period, feedback on the appropriateness of the measures and the usability of the data was gathered from clinician and clinician groups who received a report as well as the general public. Comments from field testing were summarized in a public report, which was also shared with the Clinician Expert Workgroup to consider in recommending refinements to the measures based on the testing data and feedback.

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

5.1.2.2.1 Data Provided During Field Testing

Each Field Test Report contained:

- Detailed performance results for the attributed measure, including cost measure score and breakdown of episode cost compared to the national average and TIN/TIN-NPIs with a similar patient case mix (or risk profile).
- Drill-down detail for each measure, including more detailed information on potential cost drivers in the TIN/TIN-NPI's episodes. For example:
    - Analysis of utilization and cost for the measure by the Restructured Berenson-Eggers Types of Service (BETOS) Classification System (e.g., outpatient evaluation and management services, procedures, and therapy, hospital inpatient services, emergency room services, post-acute services)[45]
    - Breakdown of costs for Part B Physician/Supplier and inpatient claims (e.g., top 5 most billed services and by risk bracket)
    - Accompanying episode-level Comma Separated Value (CSV) file with detailed information for all episodes attributed to the TIN/TIN-NPI. This file provides detailed information on every episode used to calculate your measure score, which includes winsorized observed cost, risk-adjusted cost, facilities and clinicians rendering care, the share of cost by service setting, the patient relationship code (PRC) on the trigger/reaffirming claim line.

All interested members of the public, including those who did not qualify to receive a Field Test Report, could review a series of mock reports that were representative of each measure and reporting type. Other public documentation posted during field testing included: measure specifications for each measure (comprising a Draft Cost Measure Methodology document and a Draft Measure Codes List file), a Measure Development Process document, a Frequently Asked Questions document, a Measure Testing Form (including reliability and validity data), and

[45]CMS, "Restructured BETOS Classification System https://data.cms.gov/provider-summary-by-type-of-service/provider-service-classifications/restructured-betos-classification-system

a National Summary Data Report (including national level summary statistics on the measure).[46] During field testing, Acumen conducted education and outreach activities including multiple office hours sessions with specialty societies, a publicly posted field testing webinar recording, and Quality Payment Program Help Desk support.

5.1.2.2.2 Education and Outreach

Acumen directly conducted outreach via email to tens of thousands of outreach contacts using the contact list developed through previous education and outreach and clinician engagement efforts, as well as CMS, Quality Payment Program listservs. Acumen also sent emails directly to clinicians who received the field test reports via CMS's GovDelivery.

Acumen and CMS hosted two office hours sessions in January 2022 to provide an overview of field testing to specialty societies, discuss what information their members would be particularly interested in, and answer any questions. Across both office hours sessions, there were over 35 attendees from targeted specialty societies who are likely to have members who could be attributed the measure.

Acumen worked closely with Quality Payment Program Service Center to respond to inquiries during field testing and continued to answer questions after the feedback period ended.

Acumen and CMS posted the MACRA Wave 4 Cost Measures Field Testing Webinar to the Quality Payment Program Webinar Library at the start of the field testing period.[47] The webinar recording, slides, and transcript were publicly available for review throughout field testing. The webinar presentation outlined: (i) the cost measure field testing project (ii) the measure development and re-evaluation processes, and (iii) field testing activities.

5.1.2.3 Feedback on Measure Performance and Implementation

Clinician Expert Workgroup Meetings

Feedback from the Workgroup members was recorded throughout the meeting. More formal feedback was gathered using polls, typically requesting for votes on certain specifications or appropriateness of the measure. These polls were conducted following each meeting and on an ad hoc basis, as needed.

Field Testing

In total, Acumen received 64 survey responses and 19 comment letters, including from specialty societies representing large numbers of potentially attributed clinicians.

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

5.1.2.4 Feedback from Measured Entities

Field Testing

The Field Testing Feedback Summary Report presents feedback gathered during the field testing period, including cross-measure feedback and measure-specific feedback.[48] The measure-specific feedback was used as the basis for the post-field testing refinements that

[46]The measure specifications, mock reports, Measure Development Process document, Frequently Asked Questions document, and testing 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.

[47]MACRA Wave 4 Cost Measures Field Testing Webinar materials are available on the Quality Payment Program Webinar Library: https://qpp.cms.gov/about/webinars.

[48]CMS, "2020 Field Testing Feedback Summary Report," MACRA Feedback Page, https://www.cms.gov/files/document/macra-2020-ft-feedback-summary-report.pdf.

were made to the measures. Overarching feedback about data that would be helpful for clinicians to receive was recorded and shared with CMS for future consideration. See Section 5.1.2.6 for post-field testing refinements made to the Low Back Pain measure.

5.1.2.5 Feedback from Other Users

Person and Family Engagement

Acumen incorporated thoughtful input from patients and caregivers throughout the Low Back Pain measure development process. Before each Clinical Expert Workgroup meeting, Person and Family Partners (PFPs) would provide input through focus groups and interviews to help inform the Workgroup's discussion. Attending PFPs would then present the findings for the Workgroup members, which would help shape the recommendations they made for the measure specifications. Some examples of feedback the PFP include areas for improvement as well as the types of care that were the most effective. PFPs noted that better care coordination and communication across specialists could lead to better treatment and management, especially early on. Physical therapy was listed as one of the best treatments, although PFPs recalled not having as much access to physical therapy sessions as they would like. Further, injections, acupuncture, and muscle relaxants were some of the less effective treatments.

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 the Clinician Expert Workgroup comprised of subject matter and measure-development experts. Acumen conducted analyses into potential adjustments that could be made to the measures to improve the measures' ability to assess the intended clinician population.

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 Clinician Expert Workgroup, along with the 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 Low Back Pain measure made after consideration of field testing analyses and feedback are:

- Trigger logic
    - Added codes for dry needling and remote therapeutic monitoring as trigger services
        - 20560, 20561
    - Added codes for dry needling and remote therapeutic monitoring as confirming services
        - 98975, 98977, 98980, 98981
    - Removed codes for thoracic and cervicothoracic spine diagnoses from the trigger diagnoses
    - Removed codes for non-spine specific diagnoses from the trigger diagnoses
- Measure sub-groups
    - Added history of opioid use as a risk adjustor
    - Added cognitive status/dementia as a risk adjustor
    - Added fibromyalgia as a risk adjustor
    - Added frailty proxies as a risk adjustor
    - Added laminectomy codes to the definition of the risk adjustor for patient history of prior spinal surgery

    - Added T84.84XA, T84.84XD, T84.XS: Pain Due to Internal Orthopedic Prosthetic Devices, Implants and Grafts (Initial Encounter, Subsequent Encounter, and Sequela) to the definition of the risk adjustor for patient history of prior spinal surgery
    - Excluded patients with spinal neoplasms from the measure
    - Added additional spinal infection codes to exclude from the measure
- Service assignment
    - Don't assign the costs of nutritional services to the episode group

5.2 Usability

5.2.1 Improvement

The measure has not yet been implemented, and as such has not had influence over performance.

5.2.2 Unexpected Findings

There were no unexpected findings during the development and testing of this measure. The measure has not been implemented at this time, so we don't have data that confirms unexpected findings related to its implementation.

However, Acumen did consider potential unintended consequences of having a cost measure for this clinical area (e.g., potential stinting in care to receive a better cost score). For example, the empiric validity data previously presented in Section 3.3 demonstrates that, while providing more treatment services may be associated with a worse score, it's often mediated by the cost of adverse events. In other words, attempting to stint on care will lead to an increased risk of downstream adverse events that will in turn be detrimental to the cost measure score. Therefore, it isn't in a clinician's best interest to do so to optimize their score.

Additionally, CMS monitors measures that are in use and has multiple processes in place to allow for changes to a measure if appropriate. These include i) annual maintenance for non-substantial changes and upkeep, ii) ad hoc maintenance if a specific issue occurs or a large change in clinical guidance takes place, and iii) measure reevaluation every three years where the suitability of a measure's specifications is comprehensively reassessed. If in the event the measure did have any unexpected findings, it would be identified and resolved through one of these methods.

5.2.3 Unexpected Benefits

Since the measure has not been implemented at this time, there are no testing results that identify unexpected benefits. However, many clinicians can only be assessed by the MSPB-Clinician and TPCC measures in the cost performance category currently. This measure would provide a more tailored assessment of the care they've influence over, which many clinicians may prefer to be measured by compared to the population-based cost measures like MSPB-Clinician or TPCC.

6.0 Related and Competing Measures

6.1 Relation to Other Measures

There are no competing measures with this measure. However, the following measures have been identified as potentially related.

Table 14. Quality Measures Potentially Relevant for the Low Back Pain Episode Group

Measure Title	Measure ID	Measure Description	Measure Type
Functional Outcome Assessment	CMIT 00641-C-MIPS	The percentage of visits for patients aged 18 years and older with documentation of a current functional outcome assessment using a standardized functional outcome assessment tool on the date of the encounter AND documentation of a care plan based on identified functional outcome deficiencies on the date of the identified deficiencies	Process
Functional Status Change for Patients with Low Back Impairments	CMIT 01257-C-MIPS	A patient-reported outcome measure of risk-adjusted change in functional status for patients 14 years+ with low back impairments. The change in functional status (FS) is assessed using the FOTO Low Back FS patient-reported outcome measure (PROM). The measure is adjusted to patient characteristics known to be associated with FS outcomes (risk adjusted) and used as a performance measure at the patient level, at the individual clinician level, and at the clinic level to assess quality. The measure is available as a computer adaptive test, for reduced patient burden, or a short form (static measure).	Outcome
Use of Imaging Studies for Low Back Pain	NQF 0052	The percentage of patients with a primary diagnosis of low back pain who did not have an imaging study (plain X-ray, MRI, CT scan) within 28 days of diagnosis.	Process
Back Pain After Lumbar Discectomy/Laminectomy	CMIT 05597-C-MIPS	For patients aged 18 years and older who had a lumbar discectomy/laminectomy procedure, the percentage who rate back pain as less than or equal to 3.0, or have an improvement of 5.0 points or greater on the Visual Analog Scale (VAS) Pain scale at three months (6 to 20 weeks) postoperatively.	Outcome
Back Pain After Lumbar Fusion	CMIT 05598-C-MIPS	For patients aged 18 years and older who had a lumbar fusion procedure, the percentage who rate back pain as less than or equal to 3.0, or have an improvement of 5.0 points or greater on the Visual Analog Scale (VAS) Pain scale at one year (9 to 15 months) postoperatively.	Outcome

Measure Title	Measure ID	Measure Description	Measure Type
MRI Lumbar Spine for Low Back Pain	NQF 0514	This measure evaluates the percentage of magnetic resonance imaging (MRI) of the lumbar spine studies for patients with low back pain performed in the outpatient setting where antecedent conservative therapy was not attempted prior to the MRI. Antecedent conservative therapy may include claim(s) for physical therapy in the 60 days preceding the lumbar spine MRI, claim(s) for chiropractic evaluation and manipulative treatment in the 60 days preceding the lumbar spine MRI, or claim(s) for evaluation and management at least 28 days but no later than 60 days preceding the lumbar spine MRI. The measure is calculated based on a one-year window of Medicare Claims.	Process
Leg Pain After Lumbar Fusion	CMIT 5875	For patients 18 years of age or older who had a lumbar fusion procedure, leg pain is rated by the patient as less than or equal to 3.0 OR an improvement of 5.0 points or greater on the Visual Analog Scale (VAS) Pain* scale at one year (9 to 15 months) postoperatively. *Hereafter referred to as VAS Pain.	Patient-Reported Outcome-Based Performance Measure
Functional Status After Lumbar Fusion	CMIT 5877	For patients 18 years of age and older who had a lumbar fusion procedure, functional status is rated by the patient as less than or equal to 22 OR an improvement of 30 points or greater on the Oswestry Disability Index (ODI version 2. la) at one year (9 to 15 months) postoperatively	Patient-Reported Outcome-Based Performance Measure
CAHPS 7: CAHPS for MIPS SSM: Health Status and Functional Status	CMIT 02517-C-MIPS	This is one of 10 Summary Survey Measures (SSMs) and measures patient experience of care within a group practice as part of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) for MIPS Clinician/Group Survey.	Patient Reported Outcome-Based Performance Measure)

These quality measures are relevant to the Low Back Pain episode group. For example, some measures aim to prevent negative changes in functional impairment or reduce post-procedural pain reported by specific patient cohorts who have undergone lumbar fusion or lumbar discectomy/laminectomy. One measure aims to encourage the efficient and appropriate use of imaging tests among a broad patient cohort, which has been identified as a high cost within the Low Back Pain measure. All of these measures are valuable in ensuring that cost incentives are appropriately matched with related quality incentives in this clinical area.

6.2 Harmonization

During the measure's development, the Clinician Expert Workgroup specifically considered how to align relevant cost and quality measures (e.g., episode window length). One such example was setting the attribution window to 120 days to better distinguish between efficient and inefficient care.

6.3 Competing Measures

There are no measures that conceptually address both the same measure focus and the same target population as the Low Back Pain measure.

Additional Information

Low Back Pain Clinician Expert Workgroup Members:
As noted above, the following members provided detailed feedback on the measure specifications throughout its development based on public comments, clinical expertise, and empirical analyses.

Alice Bell, PT, DPT, American Physical Therapy Association
Andrew Gordon, MD, PhD, FAAPMR, US Physiatry LLC
Carlo Milani, MD, MBA, Hospital for Special Surgery
David Seidenwurm, MD, Sutter Medical Group
Dheeraj Mahajan, MD, MBA, MPH, FACP, Chicago Internal Medicine Practice and Research (CIMPAR), SC
Erica Bisson, MD, MPH, University of Utah
Helene Fearon, BS, Fearon Physical Therapy, Inc
Jay Nathan, MD, Stanford University School of Medicine
John Heick, PT, DPT, PhD, Northern Arizona University
Kristian Anderson, DC, MS, Performance Chiropractic
Leo Bronston, DCMAppSC, Self-Employed
Luis Rodriguez, MD, FAMSSM, Baptist Health South Florida
Marcus Nynas, DC, Billings Family Chiropractic
Matthew Smith, MD, MHL, University Orthopedics
Michael Harned, MD, University of Kentucky
Michael Zychowicz, DNP, ANP, ONP, FAAN, FAANP, MSN Program, Duke University School of Nursing
Mohamad Bydon, MD, Mayo Clinic
Richard Young, MD, Acclaim Physician Group
Robert Kropp, MD, MBA, CPHI, FAAN, Aetna, Inc.
Sabrena McCarley, MBA-SL, OTR/L, CLIPP, RAC-CT, QCP, FAOTA, Select Rehabilitation
Shraddha Jatwani, MD, FACP, FACR, RhMSUS, Einstein Healthcare Network

Measure Developer Updates and Ongoing Maintenance

The measure isn't currently in use, but the earliest possible release of the measure in MIPS would be CY2025. If the measure becomes finalized for use in MIPS, it would undergo annual maintenance and a comprehensive re-evaluation every 3 years. This measure has been submitted to the 2022 Measures Under Consideration (MUC) List and may be reviewed by the MAP in winter of 2022. There are no further updates or reviews for this measure scheduled at this time.