

Quality Payment PROGRAM

Prostate Cancer

Measure Justification Form

December 2023



Table of Contents

1.0	Introduction.....	4
1.1	Project Title	4
1.2	Date	4
1.3	Project Overview	4
1.4	Measure Name.....	4
1.5	Type of Measure	4
1.6	Measure Description	4
2.0	Importance	5
2.1	Evidence to Support the Measure Focus	5
2.1.1	Logic Model.....	7
2.2	Performance Gap	7
2.2.1	Rationale.....	7
2.2.2	Performance Scores	8
2.2.3	Disparities	9
3.0	Scientific Acceptability	10
3.1	Data Sample Description	10
3.1.1	Type of Data Used for Testing.....	10
3.1.2	Specific Dataset Used for Testing	10
3.1.3	Dates of the Data Used in Testing.....	10
3.1.4	Levels of Analysis Tested	10
3.1.5	Entities Included in the Testing and Analysis	10
3.1.6	Patient Cohort Included in the Testing and Analysis	11
3.1.7	Social Risk Factors Included in Analysis	11
3.2	Reliability Testing	12
3.2.1	Level of Reliability Testing	12
3.2.2	Method of Reliability Testing.....	12
3.2.3	Statistical Results from Reliability Testing	13
3.2.4	Interpretation	14
3.3	Validity Testing	14
3.3.1	Level of Validity Testing	14
3.3.2	Method of Validity Testing	14
3.3.3	Statistical Results from Validity Testing.....	15
3.3.4	Interpretation	16
3.4	Exclusions Analysis.....	17
3.4.1	Method of Testing Exclusions.....	17
3.4.2	Statistical Results from Testing Exclusions	18
3.4.3	Interpretation	18
3.5	Risk Adjustment or Stratification	19
3.5.1	Method of Controlling for Differences	19
3.5.2	Conceptual, Clinical, and Statistical Methods.....	20
3.5.3	Conceptual Model of Impact of Social Risks	20
3.5.4	Statistical Results.....	21
3.5.5	Analyses and Interpretation in Selection of Social Risk Factors	21
3.5.6	Method for Statistical Model or Stratification Development.....	23
3.5.7	Statistical Risk Model Discrimination Statistics	23
3.5.8	Statistical Risk Model Calibration Statistics.....	24
3.5.9	Statistical Risk Model Calibration – Risk Decile	24
3.5.10	Interpretation	24
3.6	Identification of Meaningful Differences in Performance	25
3.6.1	Method	25
3.6.2	Statistical Results.....	25
3.6.3	Interpretation.....	25
3.7	Missing Data Analysis and Minimizing Bias.....	25

3.7.1	Method	25
3.7.2	Missing Data Analysis	26
3.7.3	Interpretation	26
4.0	Feasibility	27
4.1	Data Elements Generated as Byproduct of Care Processes	27
4.2	Electronic Sources	27
4.3	Data Collection Strategy	27
4.3.1	Data Collection Strategy Difficulties	27
5.0	Usability and Use	28
5.1	Use	28
5.1.1	Current and Planned Use	28
5.1.2	Feedback on the Measure by Those being Measured or Others	28
5.2	Usability	31
5.2.1	Improvement	31
5.2.2	Unexpected Findings	32
5.2.3	Unexpected Benefits	32
6.0	Related and Competing Measures	33
6.1	Relation to Other Measures	33
6.2	Harmonization	34
6.3	Competing Measures	34
	Additional Information	35

1.0 Introduction

This Measure Justification Form (MJF) provides results for the testing and evaluation of the Prostate Cancer measure. The form is intended to provide detailed information about the testing conducted on this measure, and accompanies the Measure Methodology¹ and Measure Codes List² file, which together, comprise the specifications for this cost measure.

1.1 Project Title

Physician Cost Measure and Patient Relationship Codes

1.2 Date

Information included is current on December 8, 2023

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 Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) requirements. The contract name is "Physician Cost Measure and Patient Relationship Codes (PCMP)." The contract number is 75FCMC18D0015, Task Order 75FCMC19F0004.

1.4 Measure Name

Prostate Cancer Episode-Based Cost Measure

1.5 Type of Measure

Cost/Resource Use

1.6 Measure Description

The Prostate Cancer episode-based cost measure evaluates a clinician's or clinician group's risk-adjusted and specialty-adjusted cost to Medicare for patients who receive medical care to manage and treat prostate cancer. This chronic condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during a Prostate Cancer episode.

¹CMS, "Prostate Cancer Measure Methodology," *QPP Cost Measure Information Page*, <https://www.cms.gov/medicare/quality/value-based-programs/cost-measures>

²CMS, "Prostate Cancer Measure Codes List" *QPP Cost Measure Information Page*, <https://www.cms.gov/medicare/quality/value-based-programs/cost-measures>

2.0 Importance

2.1 Evidence to Support the Measure Focus

The Prostate Cancer 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 the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA). MIPS aims to reward high-value care by measuring clinician performance through four areas: quality, improvement activities, promoting interoperability, and cost. Each category assesses different aspects of care, and the categories are weighted to combine into one composite score. CMS is introducing MIPS Value Pathways (MVPs) 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 patient costs.

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 their influence. This measure provides clinicians with information about their care costs that they can use to understand the costs associated with their decision-making. Clinicians play an essential role in the variation of healthcare expenditures due to their ability to affect costs.³ A cost measure offers an opportunity for improvement if clinicians can influence 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 through changes in clinical practice.

According to the literature and feedback received through stakeholder input activities, this measure's focus represents an area with opportunities for improvement. As discussed in the rest of this section, primary opportunities for improving prostate cancer cost outcomes include:

- I. Overtreatment among low-risk patients without limited life expectancy
- II. Overtreatment among patients with limited life expectancy,
- III. Effectiveness of prostatectomy and related adverse events,
- IV. Overuse of monitoring tests among patients under active surveillance

Prostate cancer has a relatively high survival rate, with patients more likely to die from natural causes or other diseases than from prostate cancer.⁴ While appropriate medical care is required to maintain health-related quality of life (QOL) for prostate cancer patients, more recent studies suggest that overtreatment persists in FFS⁵ and accountable care environments.⁶

Overtreatment of Medicare beneficiaries with low-risk prostate cancer can cause morbidity without survival benefit, and previous studies show that variation in treatment of low-risk

³David 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>.

⁴ Epstein MM, Edgren G, Rider JR, Mucci LA, Adami HO. Temporal trends in cause of death among Swedish and US men with prostate cancer. *J Natl Cancer Inst.* 2012;104(17):1335-1342. doi:10.1093/jnci/djs299.

⁵ Borza T, Kaufman SR, Shahinian VB, et al. Sharp Decline in Prostate Cancer Treatment Among Men in The General Population, But Not Among Diagnosed Men. *Health Aff (Millwood)*. 2017;36(1):108-115. doi:10.1377/hlthaff.2016.0739.

⁶ Modi PK, Kaufman SR, Borza T, et al. Variation in prostate cancer treatment and spending among Medicare shared savings program accountable care organizations. *Cancer*. 2018;124:3364-3371. <https://doi.org/10.1002/cncr.31573>

prostate cancer is strongly associated with provider characteristics, including medical specialty and time since graduation.⁷

Over-screening and aggressive treatments for older adults with prostate cancer are also less beneficial.^{8,9} Previous studies found no significant survival benefit within 10 years for screening in older men,^{10,11} and aggressive treatments, such as prostatectomy, do not yield survival benefits in older men and may lead to sexual dysfunctions, bowel or urinary adverse effects.^{12,13} For older patients with localized prostate cancer, clinical guidelines recommend watchful waiting, while high-risk patients are recommended to receive radiation therapy, but not surgery.²⁰

A systematic review of imaging use among patients with low-risk prostate cancer reported that patient age, comorbidities, location, and other indicators of socioeconomic status (i.e., high regional income and less income) were the most consistent determinants of overuse.¹⁴ Low-risk patients under active surveillance are recommended to receive prostate specific antigen (PSA) testing every 6 months while prostate biopsies and imaging tests, such as multiparametric MRI (mpMRI) imaging, may be done every 1 to 3 years. Monitoring tests that fall outside of these parameters should be limited.¹⁵

⁷ Hoffman KE, Niu J, Shen Y, et al. Physician variation in management of low-risk prostate cancer: a population-based cohort study. *JAMA Intern Med.* 2014;174(9):1450-1459. doi:10.1001/jamainternmed.2014.3021.

⁸ US Preventive Services Task Force, Grossman DC, Curry SJ, et al. Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement [published correction appears in JAMA. 2018 Jun 19;319(23):2443]. *JAMA.* 2018;319(18):1901-1913. doi:10.1001/jama.2018.3710.

⁹ Wolf AM, Wender RC, Etzioni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J Clin.* 2010;60(2):70-98. doi:10.3322/caac.20066.

¹⁰ Andriole GL, Crawford ED, Grubb RL 3rd, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst.* 2012;104(2):125-132. doi:10.1093/jnci/djr500.

¹¹ Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet.* 2014;384(9959):2027-2035. doi:10.1016/S0140-6736(14)60525-0.

¹² Chen RC, Clark JA, Talcott JA. Individualizing quality-of-life outcomes reporting: how localized prostate cancer treatments affect patients with different levels of baseline urinary, bowel, and sexual function. *J Clin Oncol.* 2009;27(24):3916-3922. doi:10.1200/JCO.2008.18.6486.

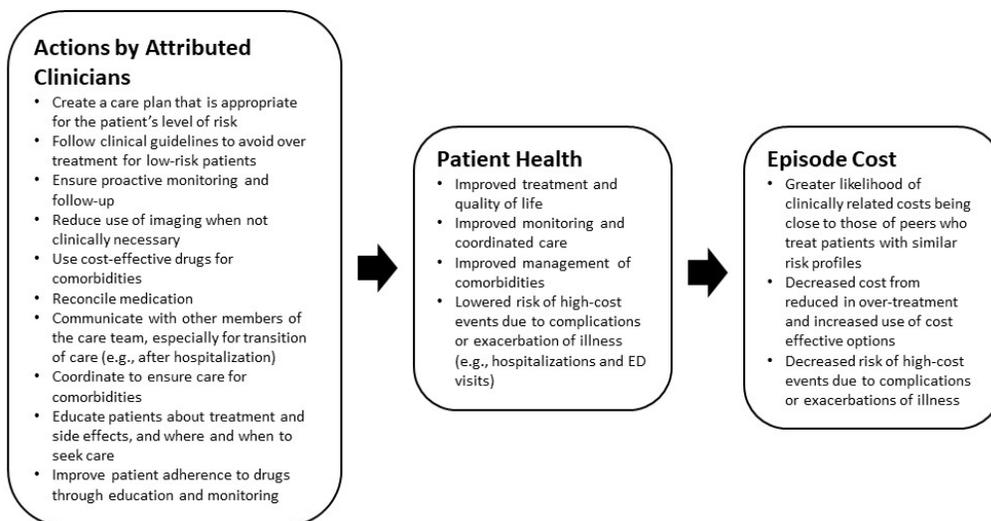
¹³ Litwin MS, Hays RD, Fink A, et al. Quality-of-life outcomes in men treated for localized prostate cancer. *JAMA.* 1995.

¹⁴ Oakes AH, Sharma R, Jackson M, Segal JB. Determinants of the overuse of imaging in low-risk prostate cancer: A systematic review. *Urol Oncol.* 2017;35(11):647-658. doi:10.1016/j.urolonc.2017.08.025.

¹⁵ Parikh NR, Chang EM, Nickols NG, et al. Cost-Effectiveness of Metastasis-Directed Therapy in Oligorecurrent Hormone-Sensitive Prostate Cancer. *Int J Radiat Oncol Biol Phys.* 2020;108(4):917-926. doi:10.1016/j.ijrobp.2020.06.009.

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

Prostate cancer is the second most common cancer among men in the United States.^{16,17} In 2019, there were over 220,000 new cases of prostate cancer reported, and more than 3.2 million men are living with this condition in the region.¹⁸ Prostate cancer is also highly prevalent among the Medicare population, as it is more likely to develop in older men. In 2020, 9.3% of male Medicare beneficiaries had a prostate cancer diagnosis, slightly increasing from 8.9% in 2017.¹⁹ This chronic condition also presents a disproportionate burden for racial or ethnic minorities and men residing in underserved and hard-to-reach geographical areas. Specifically, prostate cancer prevalence is substantially higher for Black male Medicare fee-for-service (FFS) beneficiaries than for other racial or ethnic groups (12% vs. 8%). Once diagnosed, rural-dwelling men incur more healthcare costs than their urban-dwelling counterparts.^{16,20}

Prostate cancer is costly for the Medicare population and the cost of care varies by stage of disease progression, patient risk level, and related treatment modalities. The estimated 3-year cost associated with the annual detection of this disease in older men is about \$1.2 billion.²¹ Total costs are lower in the 2-5 years after diagnosis but continue to be more expensive for

¹⁶ Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7-30. doi:10.3322/caac.21590.

¹⁷ Prostate Cancer Foundation. “Prostate Cancer Survival Rates.” *pcf.org* (2021). [Prostate Cancer Survival Rates | Prostate Cancer Foundation \(pcf.org\)](https://www.pcf.org/prostate-cancer-survival-rates)

¹⁸ “Cancer of the Prostate - Cancer Stat Facts.” SEER, seer.cancer.gov/statfacts/html/prost.html. Accessed 17 June 2022.

¹⁹ Chronic Conditions Data Warehouse. “Medicare chronic conditions charts.” *ccwdata.org* (2021). [Medicare Chronic Condition Charts - Chronic Conditions Data Warehouse \(ccwdata.org\)](https://www.cdwdata.org/medicare-chronic-conditions-charts)

²⁰ National Comprehensive Cancer Network. Guidelines for Patients: Prostate Cancer Early Stage, 2022. <https://www.nccn.org/patients/guidelines/content/PDF/prostate-early-patient.pdf>

²¹ Trogon JG, Falchook AD, Basak R, Carpenter WR, Chen RC. Total Medicare Costs Associated with Diagnosis and Treatment of Prostate Cancer in Elderly Men. *JAMA Oncol.* 2019;5(1):60-66. doi:10.1001/jamaoncol.2018.3701.

more advanced stages of cancer.²¹ Additionally, an analysis of men with localized prostate cancer enrolled in FFS found that physician treatment choice and facility factors drive cost variations more so than patient and disease characteristics, with the highest spending physicians utilizing more imaging tests, inpatient care, and radiation therapy.²² In 2020, annual prostate cancer spending was between \$18-19 billion per year, increasing faster than other cancer types.^{23,24}

The Prostate Cancer episode-based cost measure was recommended for development through feedback gathered during a public comment period. The public recommended this measure due to its high impact in terms of patient population, clinician coverage, and Medicare spending. Further, this chronic condition measure addresses a condition not captured by other episode-based cost measures in the MIPS cost performance category. Ultimately, the Prostate Cancer measure represents an opportunity to improve overall cost performance by addressing variations in costs associated with treating and managing the disease while maintaining QOL. As evidenced by the literature review, there are opportunities to improve efficiency (i.e., reduce overtreatment and overuse of select services), thereby reducing the cost to Medicare for patients receiving prostate cancer care. 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.

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).

Substantial variation is observed in the measure at the TIN and TIN-NPI levels, indicated by the interquartile ranges, standard deviations, and coefficients of variation. The 90th percentile score is more than double the 10th percentile at the TIN level and more than triple the 10th percentile at the TIN-NPI level. The results highlight an opportunity for improving clinician cost performance by closing the gap between the most and least efficient providers.

Table 1. Distribution of the Measure Score

Metric	TIN	TIN-NPI
Count	3,067	5,540
Mean Score	\$11,480	\$10,943
Score Standard Deviation	\$3,847	\$4,431
Minimum Score	\$2,244	\$1,264
Maximum Score	\$31,191	\$38,676
Score Interquartile Range (IQR)	\$4,623	\$5,617
Score Percentile		
10 th	\$6,895	\$5,837

²² Rodin D, Chien AT, Ellimoottil C, et al. Physician and facility drivers of spending variation in locoregional prostate cancer. *Cancer*. 2020;126(8):1622-1631. doi:10.1002/cncr.32719.

²³ Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML. Projections of the cost of cancer care in the United States: 2010-2020 [published correction appears in *J Natl Cancer Inst*. 2011 Apr 20;103(8):699]. *J Natl Cancer Inst*. 2011;103(2):117-128. doi:10.1093/jnci/djq495.

²⁴ Roehrig C, Miller G, Lake C, Bryant J. National health spending by medical condition, 1996-2005. *Health Aff (Millwood)*. 2009;28(2):w358-w367. doi:10.1377/hlthaff.28.2.w358.

Metric	TIN	TIN-NPI
20 th	\$8,358	\$7,220
30 th	\$9,364	\$8,327
40 th	\$10,398	\$9,273
50 th	\$11,181	\$10,333
60 th	\$12,037	\$11,438
70 th	\$12,989	\$12,696
80 th	\$14,263	\$14,252
90 th	\$16,174	\$16,769

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 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 data from the Common Working File (CWF), Long-Term Care Minimum Data Set (LTC MDS), and Medicare Enrollment Database (EDB).

3.1.2 Specific Dataset Used for Testing

The Prostate Cancer measure uses Medicare Parts A, B, and D claims data maintained by CMS. The cost measure uses Part A, B, and D claims to build episodes of care, calculate episode costs, and construct risk adjustors. These claims data are also used to designate episodes into clinically homogenous stratifications by sub-group and Part D enrollment status to ensure fair clinical comparisons among clinicians with a similar patient case mix. Episode costs are payment-standardized and risk-adjusted to compare costs across clinicians accurately. Payment standardization adjusts the allowed amount for a Medicare service to limit observed differences in costs to those that may result from healthcare 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 long-term care patients based on LTC MDS data. Specifically, the cost measure uses the LTC MDS to create the long-term care indicator variable in risk adjustment.

3.1.3 Dates of the Data Used in Testing

Prostate Cancer episodes ending from January 1, 2022, through December 31, 2022.

3.1.4 Levels of Analysis Tested

The measure was tested at the 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 Prostate Cancer measure.

Table 2: Measured Entities Demographics

Metric	TIN		TIN-NPI	
	Count	%	Count	%
Count	3,067	100.00%	5,540	100.00%
Number of Episodes Attributed	-	-	-	-
20-39 Episodes	1,000	32.61%	2,929	52.87%
40-59 Episodes	484	15.78%	1,280	23.10%
60-79 Episodes	312	10.17%	624	11.26%
80-99 Episodes	211	6.88%	327	5.90%
100-199 Episodes	509	16.60%	349	6.30%
200-299 Episodes	184	6.00%	25	0.45%
300+ Episodes	367	11.97%	6	0.11%
Census Region	-	-	-	-
Northeast	559	18.23%	1,039	18.75%

Metric	TIN		TIN-NPI	
	Count	%	Count	%
Midwest	597	19.47%	1,001	18.07%
South	1,232	40.17%	2,377	42.91%
West	670	21.85%	1,116	20.14%
Unknown	9	0.29%	7	0.13%

3.1.6 Patient Cohort Included in the Testing and Analysis

Table 3 shows the patient population for the Prostate Cancer measure testing. It consists of Medicare beneficiaries enrolled in Medicare Parts A and B who receive medical care to treat or manage prostate cancer that triggers a Prostate Cancer episode and do not meet the measure’s exclusion criteria, as outlined in section 3.4.1.

Table 3: Beneficiary Demographics

Metric	Value
Count	444,117
Mean Age	75.70 years
Male %	99.99%
Part D Enrollment %	73.42%

3.1.7 Social Risk Factors Included in Analysis

The analysis of 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.²⁵

Table 4: Social Risk Factors Available for Analysis

Variable	Advantages	Disadvantages	Used in Testing
Dual Medicare and Medicaid enrollment status	<ul style="list-style-type: none"> Available for all beneficiaries Most powerful predictor of poor outcomes²⁵ 	<ul style="list-style-type: none"> Variation in Medicaid eligibility across states 	Yes
Race/Ethnicity	<ul style="list-style-type: none"> Available for most beneficiaries, except for ambiguous categories of “Unknown” or “Other” 	<ul style="list-style-type: none"> Social risk driven by someone’s race is often correlated with and partially captured by dual status²⁵ Only 5 categories available, which may lack granularity 	No

²⁵ 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>

Variable	Advantages	Disadvantages	Used in Testing
		to fully capture disparities ^{26,27}	
ICD-10 Z codes for social determinants of health	<ul style="list-style-type: none"> • Reflects individual-level factors that influence health status and contact with health services 	<ul style="list-style-type: none"> • Not routinely and consistently coded on claims, only available for 0.1% of all fee-for-service claims in 2019²⁸ 	No
American Community Survey	<ul style="list-style-type: none"> • 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., AHRQ index, deprivation index) 	<ul style="list-style-type: none"> • 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 Prostate Cancer 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 necessary data fields used in this measure, including diagnosis and procedure codes and other elements consequential to payment. Specifically, CMS works with Zone Program Integrity Contractors, 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 under coverage, coding, and billing rules. CMS continues to perform corrective actions and give providers additional education to ensure accurate billing.

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 and Medicaid, 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>

- Lastly, to ensure claims completeness and inclusion of any corrections, the measure was developed and tested using data with three-month claims 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 how much of the variation in the measure score is explained by differences among clinician performance (i.e., signal) rather than random variation (i.e., statistical noise) among 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 2020, CMS CERT estimates that proper payment, which includes payments that met Medicare coverage, coding, and billing rules, ranged from 87.3% to 93.7% of total payments each year.²⁹ The fiscal year 2022 Medicare fee-for-service program proper payment rate was 92.5%.³⁰

Entity-level Reliability

The table below shows reliability metrics at the 20-episode testing volume threshold. While higher thresholds generally yield higher reliability results, these increases must be considered against decreasing the number of clinicians and clinician groups eligible for the measure, which would limit the applicability of measures to larger group practices and potentially limit the impact of the measure in encouraging performance improvement. For testing purposes, we used a 20-episode volume threshold. If the measure is implemented in MIPS in the future, CMS will establish a case minimum through notice-and-comment rulemaking.

²⁹Comprehensive Error Rate Testing (CERT) Program. “Appendices Medicare Fee-for-Service 2022 Improper Payments Report”. Table A6. <https://www.cms.gov/files/document/2022-medicare-fee-service-supplemental-improper-payment-data.pdf>.

³⁰Ibid.

Table 5: Reliability at the Accountability Entity Level

Reporting Level	Entities Meeting Case Minimum	Mean Reliability	Median Reliability	% Above 0.4	% Above 0.7
TIN	3,067	0.68	0.71	87.38%	52.36%
TIN-NPI	5,540	0.62	0.64	84.08%	39.26%

3.2.4 Interpretation

The results of the data element testing show very high reliability of the critical data elements used by the measure. At the entity level, the measure is moderately reliable at the TIN and TIN-NPI reporting levels, at 0.68 and 0.62, respectively. Additionally, the majority of TINs and TIN-NPIs meet or exceed the moderate reliability threshold of 0.4 at 87.38% and 84.08%, respectively. Reliability is one way to consider the extent to which performance comparisons among clinicians reflect systematic differences in performance. CMS considered existing scientific evidence on various interpretations and methods of estimating reliability. In the CY 2022 Physician Fee Schedule (86 FR 64996) rule, CMS reaffirmed the 0.4 threshold for mean reliability, continues to be appropriate for indicating moderate reliability for performance measures in the Cost category of the MIPS program.³¹

3.3 Validity Testing

3.3.1 Level of Validity Testing

The validity of the measure was tested using empirical validity at the accountable entity level (TIN and TIN-NPI).

3.3.2 Method of Validity Testing

Face Validity

The Prostate Cancer measure was developed through a structured, iterative process for gathering detailed input on the measure from recognized clinician experts. 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 between good from poor performance).

In developing this measure, Acumen incorporated input from:

- (i) a Prostate Cancer 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 [QPP Cost Measure Information 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](#).

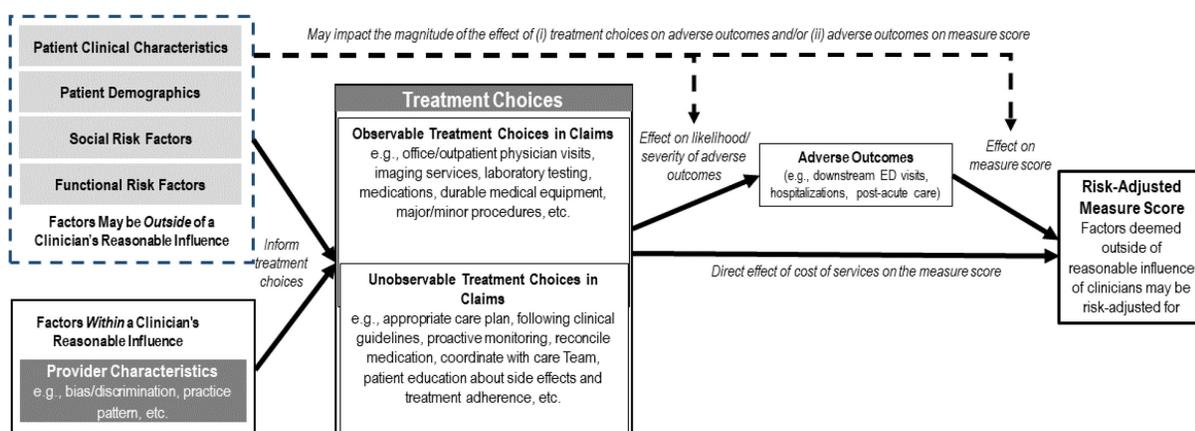
³² CMS, QPP Cost Measure Information Page, <https://www.cms.gov/medicare/quality/value-based-programs/cost-measures>

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

Empirical Validity Testing

Validity is a criterion used to assess whether the cost measure can quantify the construct it aims to measure, which is the cost directly related to treatment choices and the cost of adverse outcomes resulting from care. We evaluated the empirical validity of the Prostate Cancer 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.

Figure 2: Conceptual Model of Treatment Choices on the Measure Score



The cost measure is designed to reflect costs directly related to treatment choices and the cost of adverse outcomes resulting from care. Therefore, treatment choices, either observable in claims or otherwise, by an attributed clinician can directly impact the measure score or indirectly when they are mediated through the cost of adverse outcomes. In turn, the cost of adverse effects to the total cost captured by the measure score.

This analysis first estimates the association between treatment choices and the measure score while controlling for the cost of adverse outcomes to demonstrate that the score reflects the direct and indirect effects of treatment choices. Then, the association between treatment choices and the cost of adverse outcomes is estimated to illustrate the indirect impact.

Generally, adverse outcomes are non-trigger inpatient hospitalizations, non-trigger emergency room visits, and post-acute care. The remaining cost categories are typically considered treatment. For each category, the regression models use the mean cost across episodes 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

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 for each additional one thousand dollar of

a cost 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 dollar of a cost category that is assigned to an episode, on average, while holding the remaining categories of cost constant.

Table 6. Estimated Effect on Treatment Choices on the Measure Score

Service Categories	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	0.06 [0.04,0.07] (p < 0.01)	-	0.08 [0.07, 0.09] (p < 0.01)	-
Outpatient E/M Services	0.09 [0.04, 0.14] (p < 0.01)	0.80 [0.67,0.93] (p < 0.01)	-0.11 [-0.16, -0.07] (p < 0.01)	0.88 [0.77, 0.99] (p < 0.01)
Major Procedures	0.07 [0.06,0.08] (p < 0.01)	-0.05 [-0.07, -0.03] (p < 0.01)	0.09 [0.08,0.09] (p < 0.01)	-0.06 [-0.07, -0.05] (p < 0.01)
Ambulatory/Minor Procedures	0.16 [0.11,0.21] (p < 0.01)	-0.01 [-0.15,0.12] (p = 0.86)	0.19 [0.15,0.22] (p < 0.01)	0.09 [0.01,0.18] (p = 0.02)
Laboratory, Pathology, and Other Tests	0.14 [0.09,0.19] (p < 0.01)	0.26 [0.12,0.40] (p < 0.01)	0.14 [0.10,0.18] (p < 0.01)	0.06 [-0.04,0.16] (p = 0.23)
Imaging Services	0.03 [0.01,0.06] (p = 0.01)	-0.02 [-0.08,0.05] (p = 0.66)	0.03 [0.02,0.05] (p < 0.01)	-0.02 [-0.06,0.02] (p = 0.32)
Durable Medical Equipment and Supplies	0.32 [0.20,0.44] (p < 0.01)	0.26 [-0.06,0.59] (p = 0.11)	0.21 [0.14,0.28] (p < 0.01)	0.20 [0.03,0.37] (p = 0.02)
Anesthesia Services	2.48 [1.86,3.11] (p < 0.01)	4.00 [2.29,5.71] (p < 0.01)	1.44 [1.03,1.86] (p < 0.01)	4.33 [3.32,5.34] (p < 0.01)
Chemotherapy and Other Part B-Covered Drugs	0.00 [-0.01,0.00] (p = 0.67)	0.02 [0.01,0.04] (p < 0.01)	0.00 [0.00,0.01] (p = 0.25)	0.02 [0.01,0.03] (p < 0.01)
Part-D Drugs	0.01 [0.01,0.02] (p < 0.01)	0.01 [0.00,0.02] (p = 0.21)	0.02 [0.02,0.02] (p < 0.01)	0.01 [0.00,0.02] (p < 0.01)
All Other Services Not Otherwise Classified	0.04 [-0.03,0.10] (p=0.26)	0.07 [-0.12,0.25] (p=0.48)	0.12 [0.07,0.17] (p<0.01)	-0.05 [-0.18,0.07] (p=0.39)

3.3.4 Interpretation

Table 6 demonstrates the correlation between treatment choices and the measure score while controlling for adverse outcomes. Then the correlation between treatment choices and related adverse outcomes is calculated 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 typically considered treatment. Overall, testing results demonstrated that the cost measure reflects both the cost directly related to treatment choices and the cost of related adverse outcomes.

Model 1 shows that adverse events are associated with worse clinician performance. Outpatient evaluation/management (E/M) services are associated with worse measure performance and at the TIN-level and better measure performance at the TIN-NPI level. However, at both levels, outpatient E/Ms are associated with higher costs of adverse events, suggesting it may be possible to reduce costs associated with outpatient E/Ms. Major procedures, anesthesia, and imaging services are also associated with worse TIN and TIN-NPI scores. However, imaging services and major procedures are associated with lower costs of adverse events in Model 2, which suggests that reducing use of these services may not be beneficial. In contrast, anesthesia services are associated with higher costs of adverse events in Model 2, despite typically co-occurring with major procedures, which suggests either overuse of these services or higher usage is driven by adverse events. Similarly, laboratory testing and durable medical equipment are associated with worse scores and higher costs of adverse events in Models 1 and 2, respectively. Lastly, medications from Parts B or D are not associated with the measure score in Model 1 and have negligible association with adverse events in Model 2, suggesting the use of these services does not influence measure scores or adverse events and therefore may be candidates for improving performance.

3.4 Exclusions Analysis

3.4.1 Method of Testing Exclusions

Exclusions are used in the Prostate Cancer measure to ensure a comparable patient population within the scope of the measure's focus on patients who receive medical care to treat and manage prostate cancer 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.

- Episodes where patient death date occurred before the episode end date
 - These episodes were excluded as they may not accurately reflect a clinician's performance as the truncated episode window does not capture the full length of care intended by the measure.
- Episodes with duration less than one attribution window (i.e., shorter than 1 year)
 - These episodes were excluded because the methodology for the chronic condition measures requires at least one year of claims data to measure clinician cost performance to ensure sufficient observation of chronic care, which is often intermittent and sparse over a long period of time.
- Episodes classified as outlier cases
 - These episodes are excluded because they deviate substantially from the projected cost for a given patient risk profile.
- Episodes where there is no attributed clinician
 - These episodes are excluded because they do not have any clinicians that billed at least 30% of the clinically-related claims with a relevant diagnosis. As such, they cannot be used in the measure at the TIN-NPI level.
- Episodes with hospice care patients
 - These episodes were excluded because hospice use may be an indicator of terminal cases.

Given the rationales for these exclusions, we 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 the measure calculation to assess the distinctness between the two patient cohorts. A full list of the exclusions used for the Prostate Cancer measure is provided in the Measure Codes List available on the [QPP Cost Measure Information 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.

Table 7: Cost Statistics for Measure Exclusions

Exclusion	Episodes		Mean	Observed Cost				
	Count	% of All Episodes Meeting Triggering Logic		Percentile				
				10 th	25 th	50 th	75 th	90 th
All Episodes Meeting Triggering Logic	604,451	100.00%	\$13,617	\$361	\$703	\$2,473	\$14,035	\$36,436
Episode Length Less Than 1 Attribution Window	16,949	2.80%	\$53,708	\$1,996	\$6,477	\$26,974	\$73,398	\$139,525
Beneficiary Death in Episode	48,838	8.08%	\$46,540	\$1,699	\$5,771	\$24,416	\$65,938	\$122,521
Outlier	11,098	1.84%	\$38,142	\$1,328	\$4,122	\$36,770	\$46,488	\$136,018
No Attributed TIN-NPI	94,555	15.64%	\$16,199	\$409	\$1,007	\$5,617	\$20,552	\$38,461
TIN does not Meet Case Minimum	66,520	11.01%	\$14,538	\$419	\$742	\$2,147	\$12,664	\$38,364
TIN-NPI does not Meet Case Minimum	224,362	37.12%	\$15,063	\$386	\$699	\$2,322	\$13,938	\$42,006
Recent Hospice Use	1,065	0.18%	\$17,820	\$384	\$976	\$3,006	\$12,598	\$42,590
Reportable Episodes (if all clinicians reported as TIN at the Testing Volume Threshold)	486,931	80.56%	\$10,238	\$336	\$643	\$1,994	\$10,853	\$27,883
Reportable Episodes (if all clinicians reported as TIN-NPI at the Testing Volume Threshold)	263,217	43.55%	\$9,147	\$315	\$612	\$1,713	\$8,569	\$25,335

3.4.3 Interpretation

The statistical results show that the exclusion criteria decrease the distribution of episode cost of all episodes meeting the triggering logic, from the mean of \$13,617 to \$10,238 at the group

³³CMS, QPP Cost Measure Information Page, <https://www.cms.gov/medicare/quality/value-based-programs/cost-measures>.

reporting level and \$9,147 at the individual clinician reporting level, supporting the exclusion of these episodes to ensure a comparable patient cohort that will yield a clinically coherent measure and meaningful information to attributed clinicians. Further discussion of the results for exclusions applied based on the clinical validity of the study population are provided below.

All of the excluded episodes have higher mean observed costs than the episodes meeting the triggering logic, with the largest exclusions owing to removing episodes with no attributed clinician and applying the 20-episode testing volume threshold to ensure a sufficient sample size for the measure.

Episodes where a beneficiary died before the episode end date are excluded because they do not provide sufficient data in the episode window period. These episodes also have a higher mean observed cost than all episodes meeting triggering logic, at \$46,540 (Table 7), likely because the costs are distributed over fewer days than a typical episode.

Episodes classified as outlier cases have a mean observed episode cost of \$38,142 compared to \$13,617 for all episodes meeting triggering logic (Table 7). The wide variability of observed episode costs for outlier cases also supports their exclusion. At the 10th percentile the outlier cases observed cost is \$1,328 and at the 90th percentile the observed cost is \$136,018.

Based on testing results and input from the Prostate Cancer Clinician Expert Workgroup, episodes with hospice care patients are excluded as hospice use is likely an indicator of terminal cases. These episodes have a higher resource use pattern than all episodes meeting trigger logic, with a mean observed episode cost of \$17,820 (Table 7).

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 115 risk factors and stratification by 4 risk categories.

The risk adjustment model for the Prostate Cancer 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, age, and dual eligibility status.

The model also includes measure-specific risk factors:

- Use of androgen deprivation therapy (ADT) in the previous 1 year
- Use of chemotherapy drugs in the previous 1 year
- Use of immunotherapy drugs in the previous 1 year
- Prostatectomy in the previous 1 year
- Prostate-specific antigen (PSA) test in the previous 1 year
- Radiation therapy in the previous 1 year
- Frailty binary indicator

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

- Metastatic cancer diagnosis, metastatic drugs within one year prior
- No metastatic cancer diagnosis, metastatic drugs within one year prior

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 [QPP Cost Measure Information 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 extensively tested, we focus our testing on the adaptation of the CMS-HCC model to the Prostate Cancer measure's patient population.

The workgroup provided input on measure-specific risk adjusters 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 adjusters 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 3 shows the conceptual model that outlines how SRFs can influence the measure score, which is informed by published external research and Acumen's data analysis.^{25,35,36,37,38} 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 the influence of the attributed clinician. Risk factors, including SRFs, can influence the treatment choices and impact the size of the effect of treatment choices on mitigating the risk and cost of adverse outcomes.

A systematic approach then guides the decision of which factors to include in the risk adjustment model:

1. 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 adjusters are commonly the HCCs.
2. 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

³⁴CMS, QPP Cost Measure Information Page, <https://www.cms.gov/medicare/quality/value-based-programs/cost-measures>.

³⁵Assistant 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.

³⁶Chen 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

³⁷Medicare 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/>.

³⁸Office 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-basedpurchasing-programs>

stratified results on episode cost across many patient characteristics. We arrived at the final list of risk adjustors based on those discussions and consensus among the clinical experts.

3. During our testing phases, we also follow a structured and systematic approach to deciding whether SRFs should be adjusted for, further described in Section 3.5.5.

3.5.4 Statistical Results

The literature has extensively tested using the HCC model for Medicare claims data. Although the variables in the HCC model were selected to predict annual cost, CMS has also used this risk adjustment model in several 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 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 V24 model can be found in the Evaluation of the CMS-HCC Risk-Adjustment Model report³⁹ and the Report to Congress: Risk Adjustment in Medicare Advantage⁴⁰. 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 sub-groups.

3.5.5 Analyses and Interpretation in Selection of Social Risk Factors

To determine whether it is appropriate to risk adjust for SRFs, the following criteria are considered:

- (i) whether there is 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.

³⁹Pope, Gregory C., John Kautter, et al., "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

⁴⁰CMS, "Report to Congress: Risk Adjustment in Medicare Advantage," <https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/RTC-Dec2018.pdf>.

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

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	Metastatic Cancer Diagnosis, Metastatic Drugs 1 Year prior without Part D Enrollment	3.90%	-0.10 (p: 0.56)	-0.14 (p: 0.41)	0.02 (p: 0.94)
	Metastatic Cancer Diagnosis, Metastatic Drugs 1 Year prior with Part D Enrollment	11.82%	0.43 (p: <0.0001)	0.37 (p: <0.0001)	0.38 (p: <0.0001)
	No Metastatic Cancer Diagnosis, Metastatic Drugs 1 Year prior without Part D Enrollment	23.41%	0.27 (p: 0.01)	0.21 (p: 0.05)	0.17 (p: 0.14)
	No Metastatic Cancer Diagnosis, Metastatic Drugs 1 Year prior with Part D Enrollment	60.87%	0.24 (p: <0.0001)	0.16 (p: <0.0001)	0.16 (p: <0.0001)
TIN-NPI	Metastatic Cancer Diagnosis, Metastatic Drugs 1 Year prior without Part D Enrollment	3.90%	0.14 (p: 0.47)	0.13 (p: 0.48)	0.42 (p: 0.25)
	Metastatic Cancer Diagnosis, Metastatic Drugs 1 Year prior with Part D Enrollment	11.82%	0.44 (p: <0.0001)	0.37 (p: <0.0001)	0.37 (p: <0.0001)
	No Metastatic Cancer Diagnosis, Metastatic Drugs 1 Year prior without Part D Enrollment	23.41 %	0.29 (p: 0.01)	0.19 (p: 0.09)	0.20 (p: 0.17)
	No Metastatic Cancer Diagnosis, Metastatic Drugs 1 Year prior with Part D Enrollment	60.87%	0.24 (p: <0.0001)	0.15 (p: <0.0001)	0.15 (p: <0.0001)

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

Dual Share	TIN			TIN-NPI		
	All Episodes	Dual Episodes	Non-Dual Episodes	All Episodes	Dual Episodes	Non-Dual Episodes
All	1.13	1.32	1.12	1.16	1.31	1.15
0 – 20%	1.02	-	1.02	1.11	-	1.11
21 – 40%	1.11	1.33	1.10	1.10	1.26	1.10
41 – 60%	1.15	1.25	1.14	1.18	1.32	1.18
61 – 80%	1.15	1.31	1.14	1.15	1.29	1.14
81 – 100%	1.23	1.39	1.20	1.22	1.32	1.21

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

Reporting Level	Significantly Worse	Equally Well	Significantly Better
TIN	9.09%	90.55%	0.37%
TIN-NPI	7.74%	92.03%	0.23%

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	67.33%	5.87%
TIN-NPI	59.86%	4.01%

The results suggest that there is a statistically significant association between the patient's dual status and episode cost across the different sub-groups, as observed in Table 8. This association is relatively stable and remains statistically significant after adding variables to account for clinician-level factors, which suggests that the patient-level factors are more influential than clinician-level factors. Performance degradation is associated with higher share of dual beneficiaries and is degradation observed on both dual and non-dual episodes when stratified (Table 9). While many clinicians are able to perform equally well on their dual episodes and non-dual episodes, there are still a substantial number of clinicians performing significantly worse on their dual episodes than their non-dual episodes, which suggests that clinicians aren't able to fully mitigate the effect of SRFs (Table 10). Lastly, risk adjusting for dual status appears to change the performance ranking for a subset of clinicians (Table 11).

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 Prostate Cancer cost measure, calculated by dividing explained sum of squares by total sum of squares is 0.32. The adjusted R-squared is also 0.32. More

information on discrimination testing for the CMS-HCC model can be found at Pope et al. 2011.⁴¹

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

Across all risk deciles, the predictive ratio is 1.00. Analysis of predictive ratios by risk decile for the measure shows moderate variation among risk deciles, as predictive ratios range from 0.87 to 1.09 across all risk deciles.

Table 12: Predictive Ratio by Decile of Predicted Episode Cost

Decile	Average Predictive Ratio
Decile 1	0.87
Decile 2	0.93
Decile 3	0.95
Decile 4	0.96
Decile 5	1.00
Decile 6	0.96
Decile 7	0.88
Decile 8	0.92
Decile 9	0.97
Decile 10	1.09

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.⁴² 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 are not adjusted for by the measure, such as the clinician's performance. The objective of a cost measure is to evaluate and differentiate the performance of clinicians. Therefore, achieving high explained variance is optional because the measure should only adjust for some variations in the cost of care. In collaboration with the experts from our clinical workgroup, this measure only adjusts for factors that are deemed outside the reasonable influence of clinicians. The service assignment rules provide context for which costs are included in the measure and which are not.

Table 12 shows that the risk adjustment model shows moderate variation across risk deciles, with the range between 0.87 and 1.09 across risk deciles. For most risk deciles, the average predictive ratio is close to 1.00.

⁴¹Pope, Gregory C., John Kautter, et al., "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

⁴²Pope, Gregory C., John Kautter, Melvin J. Ingber, Sara Freeman, Rishi Sekar, and Cordon Newhart. "Evaluation of the CMS-HCC Risk-Adjustment Model: Final Report." RTI International: March 2011.

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 adverse 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. There is a difference in mean score for TIN and TIN-NPI levels because each level has its own attribution rules, which resulted in slightly different populations of episodes used for measure score calculation (Table 1). However, clinicians are only compared to their peers at either the TIN or TIN-NPI level, therefore the differences in score across different levels can be ignored. Additionally, the testing results show that 11.98% and 10.66% of episodes had a clinically related emergency department visit and anesthesia services, respectively, with mean observed episode costs of \$20,976 and \$29,493. Further, episodes with clinically related major procedures and inpatient hospitalizations also had high mean observed costs of \$27,347 and \$32,457.

3.6.3 Interpretation

There is 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. Furthermore, there are also opportunities to reduce costs associated with adverse events, such as emergency department visits, major procedures, and inpatient hospitalizations. Episodes with clinically related emergency department visits, major procedures, or inpatient hospitalizations are costly to Medicare with mean observed costs (\$20,976, \$27,347, \$33,792) much higher than the average Prostate Cancer episode with a mean risk-adjusted cost of \$10,566.

3.7 Missing Data Analysis and Minimizing Bias

3.7.1 Method

Since CMS uses Medicare claims data to calculate the Prostate Cancer measure, Acumen expects a high degree of data completeness. To further ensure that we have complete and accurate data for each patient, Acumen excludes episodes where patient date of birth information (an input to the risk adjustment model) cannot 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 Prostate Cancer 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 Prostate Cancer 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 in the measure reporting.

As a note, the episode counts below reflect exclusion from the initial population of triggered episodes. After the missing data exclusions are applied, we 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	Episode Count	Observed Cost					
		Mean	Percentile				
			10 th	25 th	50 th	75 th	90 th
All Episodes	912,796	\$13,436	\$336	\$671	\$2,275	\$13,175	\$35,655
Beneficiary Resides Outside U.S. or its Territories	665	\$14,250	\$379	\$916	\$3,550	\$15,483	\$36,829
Primary Payer Other Than Medicare	72,233	\$12,746	\$316	\$669	\$2,199	\$11,903	\$32,674
No Continuous Enrollment in Medicare Parts A and B, and Any Enrollment in Part C	71,849	\$12,747	\$175	\$436	\$1,563	\$11,159	\$33,312

3.7.3 Interpretation

The results show that the episodes with missing data are not substantially different than all episodes in the initial population in terms of cost (Table 13). It is appropriate to remove these episodes as they are likely indicators of a discontinuation of the patient-clinician relationship or an absence of Medicare usage, and therefore do not provide sufficient data during the episode window. Given their limited frequencies, the impact of removing these episodes on the overall measure should be minimal while ensuring that clinicians are fairly evaluated on episodes with complete information.

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 CWF maintained at the CMS Baltimore Data Center. Healthcare providers submit Medicare claims to a Medicare Administrative Contractor (MAC), which 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 such, it is not practical to wait until all claims for a given month are finalized before calculating the measure, resulting in a trade-off between efficiency (accessing the data on time) and accuracy (waiting until most claims are finalized) when determining the duration (i.e., the “claims run-out” period) after which to pull claims data. To determine the appropriate claims run-out period, Acumen has tested 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 CMS adopts this measure for use in a program, calculation and reporting would align with the program’s reporting practices.

4.3.1.2 Missing Data

This measure requires complete beneficiary information, therefore, a small number of episodes with missing data are excluded to ensure data completeness and accurate comparability across episodes. For example, episodes where the beneficiary was not enrolled in Medicare Parts A and B for the 120 days before the episode start date are excluded from this measure. Excluding these episodes 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 cannot be located are excluded from the measure.

4.3.1.3 Sampling

During measure testing, Acumen noted that episodes in which the beneficiary died before the episode end date exhibited different cost distributions than other episodes. As such, this measure excludes episodes to avoid negatively impacting clinician scores.

5.0 Usability and Use

5.1 Use

5.1.1 Current and Planned Use

The measure is not currently in use but is intended for use in a payment program and could eventually be publicly reported. It 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 CMS to approve this measure for use in MIPS, it must be reviewed by the Pre-Rulemaking Measure Review process (PRMR; formerly referred to as the Measure Application Partnership [MAP]) and then undergo the notice-and-comment process. Given these next steps, the earliest the measure could be used in MIPS is CY 2025. 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 Prostate Cancer measure development, we used an iterative and extensive process to gather feedback on the measure and its results to ensure that it can be used appropriately in the MIPS program by clinicians and clinician groups who practice in this clinical area. This process also seeks to ensure that the measured entities can understand and interpret their performance results to help support decision-making. A couple of the main ways we gathered input was through reoccurring Clinician Expert Workgroup meetings, which incorporated feedback from the patient and caregiver perspective, empirical data, and discussion between clinician experts who recommend measure specifications, and through the national field testing of the measures.

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) to inform the Clinician Expert Workgroup members' recommendations. 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 is appropriately assessing costs for these clinicians.

Field Testing

Additionally, Acumen and CMS nationally field tested the draft Prostate Cancer measure, along with 4 other episode-based cost measures, for a 4-week comment period (January 17 to February 14, 2023). We provided a Field Test Report with performance data to all clinician groups and clinicians who were attributed 20 or more episodes, which was the testing volume threshold.⁴³ 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 stakeholders as possible. A total of 9,206 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.

⁴³The 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 before or during each of the Clinician Expert Workgroup Meetings: The Workgroup Webinar, Service Assignment and Refinement Webinar, and Post-Field Test Refinement Webinar. 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 appropriately assessed 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, the measured entities (i.e., MIPS-eligible clinicians and clinician groups who received a report) and the general public provided feedback on the appropriateness of the measures and the usability of the data. The public comments were summarized in a report, which was shared with the Clinician Expert Workgroup for consideration when recommending refinements to the measures based on the testing data and feedback.

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

Data Provided During Field Testing

Each Field Test Report contained:

- 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 BETOS Classification System (e.g., outpatient evaluation and management services, procedures, and therapy, hospital inpatient services, emergency room services, post-acute services)⁴⁴
 - 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 stakeholders, 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,

⁴⁴CMS, "Restructured BETOS Classification System <https://data.cms.gov/provider-summary-by-type-of-service/provider-service-classifications/restructured-betos-classification-system>

and a Measure Testing Form (including reliability and validity data).⁴⁵ During field testing, Acumen conducted education and outreach activities for interested parties, including multiple office hours sessions with specialty societies, a publicly posted field testing webinar recording, and Quality Payment Program Help Desk support.

Education and Outreach

Acumen directly conducted outreach via email to tens of thousands of interested parties using a contact list developed through previous public engagement efforts, as well as CMS and Quality Payment Program (QPP) listservs. Acumen also emailed clinicians who received the field test reports via CMS's GovDelivery.

Acumen and CMS hosted two office hours sessions in January 2023 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 attendees from targeted specialty societies who are likely to have members who could be attributed the measure.

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

Acumen and CMS hosted the public 2023 MACRA Cost Measures Field Testing webinar in January 2023, where interested parties could learn more about field testing and the measures.⁴⁶ The webinar presentation outlined: (i) the cost measure field testing project (ii) the measure development and re-evaluation processes, and (iii) field testing activities. There was also an opportunity to ask questions during the Q&A portion of the webinar. The webinar recording, slides, and transcript were then made available for the public to review.

5.1.2.3 Feedback on Measure Performance and Implementation

Clinician Expert Workgroup Meetings

Feedback from the Workgroup members were 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 48 survey responses and 5 comment letters, including from specialty societies representing large numbers of potentially attributed clinicians and from persons with lived experiences.

Survey responses and comment letters were collected via two online surveys, 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.⁴⁷ The

⁴⁵The measure specifications, mock reports, Measure Development Process document, Frequently Asked Questions document, and testing documents are posted on the Cost Measures Information Page:

<https://www.cms.gov/medicare/quality/value-based-programs/cost-measures>.

⁴⁶MACRA Wave 4 Cost Measures Field Testing Webinar materials are available on the Quality Payment Program Webinar Library: <https://qpp.cms.gov/about/webinars>.

⁴⁷CMS, "2023 Field Testing Feedback Summary Report," Cost Measures Information Page, <https://www.cms.gov/files/document/field-testing-feedback-summary-report-23-wave-5.pdf>.

measure-specific feedback was used as the basis for the post-field testing refinements that 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 Prostate Cancer measure.

5.1.2.5 Feedback from Other Users

Person and Family Engagement

Acumen incorporated thoughtful input from patients and caregivers throughout the Prostate Cancer measure development process. Before each Clinician 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 the types of services that patients with prostate cancer typically receive (e.g., biopsies, imaging services, radiation therapy, surgery, chemotherapy and ADT) and the multiple specialties involved in their care (e.g., urologists, nurse practitioners, radiation/medical oncologists, and hematologists). PFPs also noted the challenges they face when receiving care, including limited patient-provider communication, lack of care coordination, inadequate pain management, and lack of education on the different treatment modalities and their side effects. Moreover, they emphasized the importance of timely diagnosis and treatment in managing disease severity and the need for psychosocial support for patients undergoing treatment for prostate cancer.

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 their ability to assess the intended clinician population.

After field testing, Acumen compiled the feedback provided through the surveys 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 Prostate Cancer measure made after consideration of field-testing analyses and stakeholder feedback are:

- Risk Adjustment
 - Extended the length of the lookback window from 120 days to 365 days for identifying treatment modalities that a patient received prior to the start of the episode
 - Added the following risk adjustor variables: radiation therapy in the previous 1 year, use of ADT in previous 1 year, prostatectomy in the previous 1 year, PSA testing in the previous 1 year

5.2 Usability

5.2.1 Improvement

The measure has not yet been implemented, and as such has not had influence over performance. Our testing suggests that there is a sufficiently large difference in measure scores among clinicians to meaningfully determine a difference in performance. The potential for this

measure to distinguish between good and poor performance is promising in its ability to encourage improvement in cost efficient care.

Additionally, the face validity results suggest that the Clinician Expert Workgroup believes the measure assess care within the influence of the clinician and can positively impact care provision and coordination.

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 do not have data that confirms unexpected findings related to its implementation. However, Acumen considered the 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 instance, the empiric validity data previously presented in section 3.3 demonstrates that while medications from Part B or D may be costly, they are not a major driver of the measure score, therefore, demonstrating the robustness of the risk adjustment model and the ability of the cost measure to differentiate performance that is most relevant to the treatment and management of patients with prostate cancer.

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 have 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 Prostate Cancer Measure

Measure Title	CMIT Measure ID	Measure Description	Measure Type
Avoidance of Overuse of Bone Scan for Staging Low Risk Prostate Cancer Patients	00614	Percentage of patients, regardless of age, with a diagnosis of prostate cancer at low (or very low) risk of recurrence receiving interstitial prostate brachytherapy, OR external beam radiotherapy to the prostate, OR radical prostatectomy who did not have a bone scan performed at any time since diagnosis of prostate cancer.	Process
Combination Androgen Deprivation Therapy (ADT) for High Risk or Very High Risk Prostate Cancer	00615	Percentage of patients, regardless of age, with a diagnosis of prostate cancer at high or very high risk of recurrence receiving external beam radiotherapy to the prostate who were prescribed androgen deprivation therapy in combination with external beam radiotherapy to the prostate.	Process
Radical Prostatectomy Pathology Reporting	00623	Percentage of radical prostatectomy pathology reports that include the pT category, the pN category, the Gleason score and a statement about margin status.	Process
Hospital-Wide, 30-Day, All-Cause Unplanned Readmission (HWR) Rate	00356	This measure is a re-specified version of the measure, "Risk-adjusted readmission rate (RARR) of unplanned readmission within 30 days of hospital discharge for any condition" (NQF 1789), which was developed for patients 65 years and older using Medicare claims. This re-specified measure attributes outcomes to MIPS participating clinician groups and assesses each group's readmission rate. The measure comprises a single summary score, derived from the results of five models, one for each of the following specialty cohorts (groups of discharge condition categories or procedure categories): medicine, surgery/gynecology, cardio-respiratory, cardiovascular, and neurology.	Outcome
Bone Density Evaluation for Patients with Prostate Cancer and Receiving Androgen Deprivation Therapy	00091	Patients determined as having prostate cancer who are currently starting or undergoing androgen deprivation therapy (ADT), for an anticipated period of 12 months or greater and who receive an initial bone density evaluation. The bone density evaluation must be prior to the start of ADT or within 3 months of the start of ADT.	Process

Measure Title	CMIT Measure ID	Measure Description	Measure Type
Nuclear Medicine: Correlation with Existing Imaging Studies for All Patients Undergoing Bone Scintigraphy	00470	Percentage of final reports for all patients, regardless of age, undergoing bone scintigraphy that include physician documentation of correlation with existing relevant imaging studies (e.g., x-ray, Magnetic Resonance Imaging (MRI), Computed Tomography (CT), etc.) that were performed.	Process
Oncology: Medical and Radiation – Plan of Care for Pain	00473	Percentage of visits for patients, regardless of age, with a diagnosis of cancer currently receiving chemotherapy or radiation therapy who report having pain with a documented plan of care to address pain.	Process
Oncology: Medical and Radiation – Pain Intensity Quantified	00474	Percentage of patient visits, regardless of patient age, with a diagnosis of cancer currently receiving chemotherapy or radiation therapy in which pain intensity is quantified.	Process
Surgical Site Infection (SSI)	00001	Percentage of patients aged 18 years and older who had a surgical site infection (SSI).	Outcome

The MIPS quality measures listed above are related to the Prostate Cancer measure because they may include metrics focused on similar patient cohorts or clinically related to the care provided for the episode group. While four quality measures are specific to prostate cancer care, the remaining measures apply to a broader cohort of patients with pain management related to radiation therapy or chemotherapy, imaging services, surgical site infection, or other treatment-related complications.

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). The Prostate Cancer measure has the potential to be used in the Advancing Cancer Care MVP. MVPs offer a participation framework meant to align cost and quality measures providing a degree of standardization to hold clinicians accountable for their clinical decisions in a consistent manner. MVPs also seek to connect measures with improvement activities to the relevant area of clinical practice. While there are no improvement activities in MIPS that are specific to the prostate cancer clinical area. However, there is an improvement activity related to chronic care, Chronic Care and Preventative Care Management for Empaneled Patients (IA_PM_13), which may correlate with the Prostate Cancer measure as it aims to improve outcomes for patients that have chronic conditions or diseases.

6.3 Competing Measures

There are no measures that conceptually address both the same measure focus and the same target population as the Prostate Cancer measure.

Additional Information

Prostate Cancer 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.

Daniel Barocas, MD, MPH, Society of Urologic Oncology
Robert Dreicer, MD, MS, FASCO, MACP, American Society of Clinical Oncology
Sarah Eakin, MD, College of American Pathologists
Andrew Harris, MD, American Urological Association
Karen Henry, DNP, MSN, APRN, FNP-BC, AOCNP, Advanced Practitioner Society of Hematology and Oncology
John Lam, MD, MBA, FACS, American Urological Association
Kathleen Latino, MD, American Urological Association
John Lin, MD, American Society of Clinical Oncology
Join Luh, MD, American Society for Radiation
Megan May, PharmD, BCOP, Advanced Practitioner Society of Hematology and Oncology
Timothy McClure, MD, Society of Interventional Radiology
Alicia Morgans, MD, MPH, American Society of Clinical Oncology
Christopher Peters, MD, American Board of Radiology – Radiation Oncology
David Seidenwurm, MD, American College of Radiology
Abhishek Solanki, MD, MS, American Society for Radiation Oncology
Barbara Spivak, MD, American Medical Association
Sean Woolen, MD, MSc., American College of Radiation

Measure Developer Updates and Ongoing Maintenance

The measure is not 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 is included on the 2023 Measures Under Consideration (MUC) List and will be reviewed by PRMR in winter of 2023-2024. There are no further updates or reviews for this measure scheduled at this time.