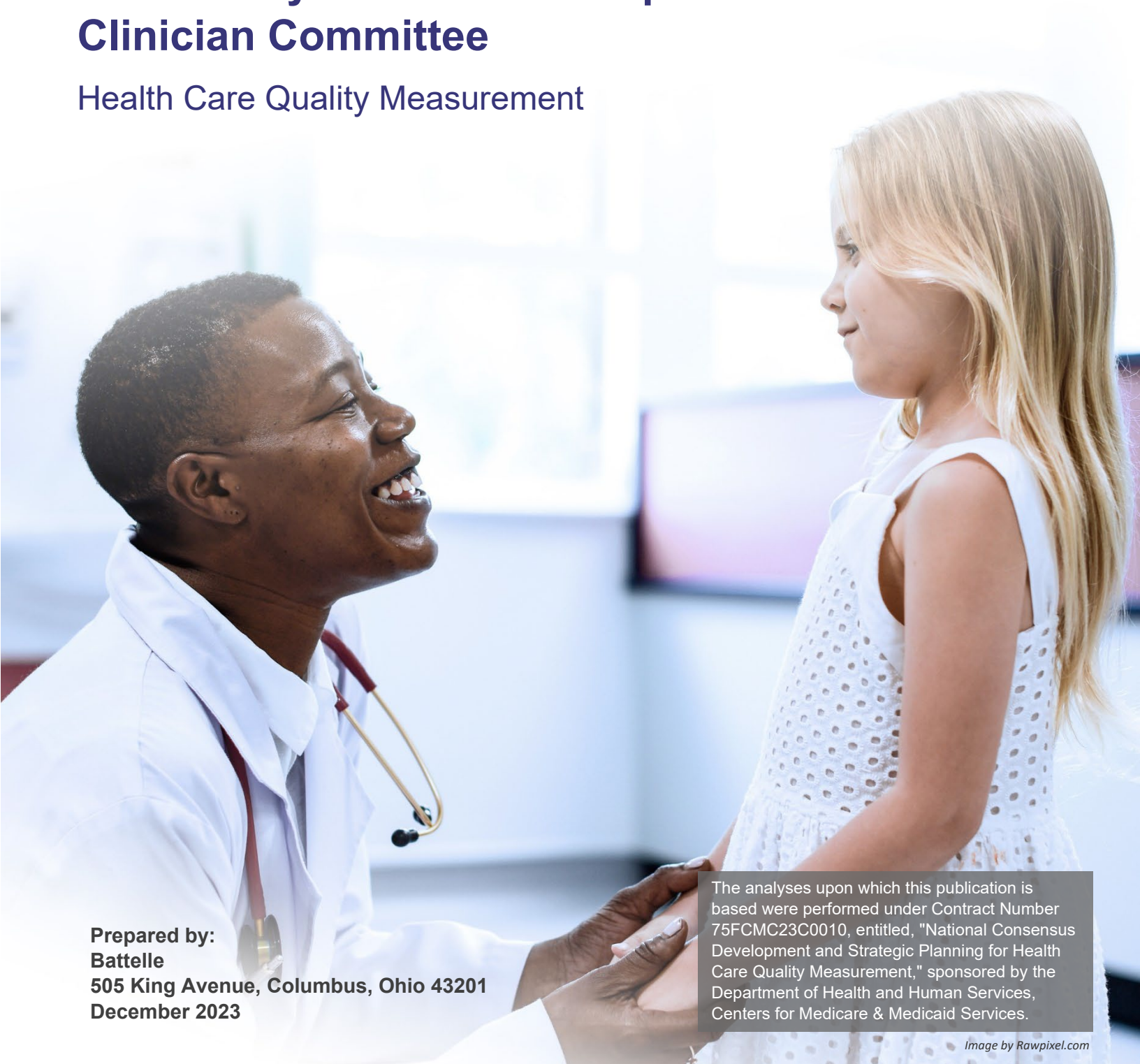
Partnership for Quality Measurement
Powered by Battelle

National Consensus Development and Strategic Planning for

2023 Pre-Rulemaking Measure Review (PRMR) Preliminary Assessment Report: Clinician Committee

Health Care Quality Measurement



Prepared by:
Battelle
505 King Avenue, Columbus, Ohio 43201
December 2023

The analyses upon which this publication is based were performed under Contract Number 75FCMC23C0010, entitled, "National Consensus Development and Strategic Planning for Health Care Quality Measurement," sponsored by the Department of Health and Human Services, Centers for Medicare & Medicaid Services.

Image by Rawpixel.com

Table of Contents

Page

List of Tables

List of Figures

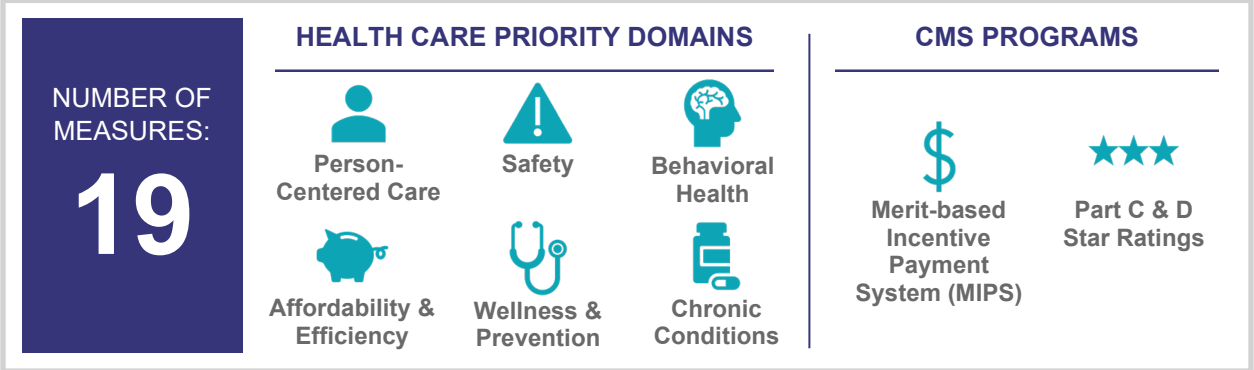
Executive Summary

The Pre-Rulemaking Measure Review (PRMR) process, undertaken yearly, informs the selection of health care quality and efficiency measures for use in Centers for Medicare & Medicaid Services (CMS) Medicare quality programs. Each cycle begins with the publication of the Measures Under Consideration (MUC) list. The MUC list is reviewed by interested parties, selected to serve on PRMR committees. The PRMR process engages a diverse group of interested parties in making consensus-based recommendations regarding the inclusion of considered measures.

This PRMR Preliminary Assessment Report for the Clinician Committee provides PRMR Advisory and Recommendation Group members with a detailed baseline evaluation of the measures under consideration for Clinician-relevant CMS programs this PRMR cycle. The findings of this report will enable committee members to further examine and discuss measure suitability for the selected CMS program(s) during the PRMR Recommendation Group Meetings in January 2024.

Measure assessments included evaluation of submission materials such as CMS MUC Entry/Review Information Tool (MERIT) submission forms, reliability and validity testing results, and summaries of evidence for measure relevance to specific program populations. A team of Battelle measure evaluators reviewed submission materials for the 19 measures under consideration for the Merit-based Incentive Payment System (MIPS) and the Part C and D Star Ratings Program and applied standardized criteria across the domain of meaningfulness, including elements such as importance, conformance, feasibility, validity, reliability, and usability (Figure 1). The measure evaluations and descriptions of available evidence in this report will inform PRMR committee consideration of measure meaningfulness as well as additional criteria of appropriateness of scale and time to value realization during later stages of the PRMR cycle.

Figure 1. Clinician Committee Measures Under Consideration



Chapter 1. Introduction

1.1 Overview

The goal of the PRMR process is to inform the selection of health care quality and efficiency measures for use in CMS Medicare quality programs. Input from interested parties informs these recommendations throughout the measure life cycle. The cornerstone of a transparent and inclusive consensus-based process is effective engagement of interested parties. This ensures that meaningful feedback is provided to CMS on all measures proposed for inclusion in CMS quality programs. The interested parties include those who are impacted or affected by the use of quality and efficiency measures. Interested parties come from a variety of places (Figure 2) and represent a diverse group of people.

Figure 2. PRMR Interested Parties

The Health & Human Services (HHS), per statute[1], publishes annually (by December 1) a list of measures under consideration (MUC) for future federal rulemaking. The PRMR process makes consensus recommendations regarding the inclusion of measures being considered for CMS quality reporting and value-based programs. PRMR's review focuses on a measure's appropriateness for a specific program. It assesses if, within the proposed program, the measure is meaningful, tailored to the program's unique needs, balanced, and scaled to meet program-specific goals, and demonstrates a clear vision of near- and long-term program impacts.

Previously conducted via the Measure Applications Partnership (MAP) process, the annual review of measures under consideration is now called Pre-Rulemaking Measure Review (PRMR, pronounced Primer).

[1] Section 3014 of the Patient Protection and Affordable Care Act of 2010 (ACA) (P.L. 111-148) created section 1890A of the Social Security Act (the Act), which required HHS to establish a federal pre-rulemaking process for the selection of quality and efficiency measures for use by HHS.

1.2 Relevant CMS Programs

The Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) ended the Sustainable Growth Rate (SGR) formula, which would have resulted in a significant cut to payment rates for clinicians participating in Medicare. MACRA requires CMS to implement an incentive program for clinicians.[2] This program, referred to as the Quality Payment Program, provides two participation pathways for clinicians: the Merit-based Incentive Payment System (MIPS): Traditional MIPS or MIPS Value Pathways (MVPs) and the Advanced Alternative Payment Models (Advanced APMs).

MIPS combines three Medicare "legacy" programs—the Physician Quality Reporting System (PQRS), Value-based Payment Modifier (VM), and the Medicare EHR (Electronic Health Record) Incentive Program for Eligible Professionals—into a single program. Under MIPS, four connected performance categories affect a clinician's future Medicare payments: Quality, Promoting Interoperability, Improvement Activities, and Cost. The performance categories included in the Clinician Committee this year are Quality and Cost. The Quality performance category assesses the quality of care delivered, as evidenced by the performance on quality measures. The Cost performance category assesses the relative cost of the care provided based on Medicare claims. Cost measures are risk adjusted and only include clinically related services and apply certain exclusions to ensure only relevant costs are measured.

The Part C and D Star Ratings Program is a system that measures the quality of health and drug services received by consumers enrolled in Medicare Advantage (MA) and Prescription Drug Plans (PDPs or Part D plans). This rating system helps beneficiaries compare the quality of health and drug plans they are offered and empowers them to make choices regarding their health care. The sources used to rate the plans are data collected by CMS contractors, CMS administrative data, survey of enrollees, and data from the health and drug plans. Plans are ranked on a five-star scale, one star being the lowest and five stars being the highest.

The Medicare Shared Savings Program (Shared Savings Program) is Medicare's national value-based payment program for Accountable Care Organizations (ACO). ACOs facilitate coordination and cooperation among health care providers to improve the quality of care for Medicare Fee-For-Service (FFS) beneficiaries and reduce the rate of growth in health care costs.

More information on both programs and structure for 2023 is available in Appendix A, which includes excerpts from the CMS Measures Under Consideration List: Program-Specific Measure Needs and Priorities.

1.3 Measures Under Consideration

For the 2023 PRMR review cycle, 19 measures are under consideration for inclusion in the Merit-based Incentive Payment System – Cost, Merit-based Incentive Payment System – Quality, and the Part C and D Star Ratings [Medicare] programs. Table 1 lists the measures under consideration for review by the Clinician committee and their associated CMS Cascade of

[2] CMS. 2023 MUC List Program Specific Needs and Priorities. Accessed 8 November 2023. https://mmshub.cms.gov/sites/default/files/2023-MUC-List-Program-Specific-Measure-Needs-and-Priorities.pdf

Meaningful Measures priority area.[3] "Cascade Priority" is included to show the alignment of each measure with a meaningful measure area and to provide more context for what the measure's addition could bring to the selected CMS program. These measures are available for public comment at the PQM website December 1-22, 2023.

Table 1.3.1. MUC List Measures by Cascade of Meaningful Measures Priority

MUC ID	Measure Title	Cascade Priority	CMS Program
MUC2023-137	Initial Opioid Prescribing for Long Duration (IOP-LD)	Safety	Part C & D Star Ratings
MUC2023-141	Positive PD-L1 Biomarker Expression Test Result Prior to First-Line Immune Checkpoint Inhibitor Therapy	Chronic Conditions	Merit-based Incentive Payment System-Quality
MUC2023-161	Appropriate Germline Testing for Ovarian Cancer Patients	Chronic Conditions	Merit-based Incentive Payment System-Quality
MUC2023-162	Patient-Reported Pain Interference Following Chemotherapy among Adults with Breast Cancer	Person-Centered Care	Merit-based Incentive Payment System-Quality
MUC2023-164	Adult COVID-19 Vaccination Status	Wellness and Prevention	Merit-based Incentive Payment System-Quality
MUC2023-179	Initiation and Engagement of Substance Use Disorder Treatment (IET)	Behavioral Health	Part C & D Star Ratings
MUC2023-190	Patient-Reported Fatigue Following Chemotherapy among Adults with Breast Cancer	Person-Centered Care	Merit-based Incentive Payment System-Quality
MUC2023-199	Connection to Community Service Provider	Equity	Medicare Shared Savings Program
MUC2023-201	Cataract Removal with Intraocular Lens (IOL) Implantation	Affordability and Efficiency	Merit-based Incentive Payment System-Cost
MUC2023-203	Chronic Kidney Disease	Affordability and Efficiency	Merit-based Incentive Payment System-Cost
MUC2023-204	End-Stage Renal Disease	Affordability and Efficiency	Merit-based Incentive Payment System-Cost
MUC2023-205	Inpatient (IP) Percutaneous Coronary Intervention (PCI)	Affordability and Efficiency	Merit-based Incentive Payment System-Cost
MUC2023-206	Kidney Transplant Management	Affordability and Efficiency	Merit-based Incentive Payment System-Cost
MUC2023-207	Prostate Cancer	Affordability and Efficiency	Merit-based Incentive Payment System-Cost

[3] CMS. Cascade of Meaningful Measures. Accessed 6 November 2023. https://www.cms.gov/cascade-measures#:~:text=The%20Cascade%20of%20Meaningful%20Measures,may%20need%20to%20be%20developed.

MUC ID	Measure Title	Cascade Priority	CMS Program
MUC2023-208	Respiratory Infection Hospitalization	Affordability and Efficiency	Merit-based Incentive Payment System-Cost
MUC2023-209	Rheumatoid Arthritis	Affordability and Efficiency	Merit-based Incentive Payment System-Cost
MUC2023-211	Melanoma: Tracking and Evaluation of Recurrence	Wellness and Prevention	Merit-based Incentive Payment System-Quality
MUC2023-212	Level I Denials Upheld Rate Measure	Person-Centered Care	Part C & D Star Ratings
MUC2023-210	Resolution of At Least 1 Health-Related Social Need	Equity	Medicare Shared Savings Program

Chapter 2. Preliminary Assessment Methodology

2.1 Goals and Objectives

The goal of this PRMR Preliminary Assessment Report for the Clinician Committee is to provide committee members with a thorough and standardized baseline evaluation of the measures under consideration for Clinician programs. This preliminary assessment supports committee members as they further examine and discuss measure suitability to the selected CMS program before and during the PRMR Recommendation Group Meetings.

To achieve this goal, Battelle staff conducted preliminary assessments of each measure with three objectives in mind:

1) To assess completeness of measure information provided in the CMS MUC Entry/Review Information Tool (CMS MERIT) submission and review available testing/performance data.
2) To evaluate measures against consistent criteria with an emphasis on importance, conformance, feasibility, reliability, validity, and usability (i.e., meaningfulness).
3) To provide a summary of findings based on the evaluation criteria that describes the likelihood that each measure meets “meaningfulness” requirements for use in a CMS program. Note: Measures that have received Consensus-Based Endorsement are assumed to largely meet the meaningfulness criteria, although reviewers are asked to consider the specific needs of the selected program when evaluating this for PRMR.

2.2 Data Sources

To conduct this preliminary assessment, Battelle staff reviewed submission documentation provided in the CMS MERIT system. The types of information provided varied by measure but generally fell into the following categories: CMS MERIT Submission Form, Measure Information Forms, peer-reviewed literature, clinical practice guidelines, validity and reliability testing methods and results, and electronic clinical quality measure (eCQM) feasibility testing information, if applicable.

2.3 Evaluation Criteria

A team of experienced measure evaluators reviewed the available information for each measure from the data sources listed above and compared it against evaluation criteria for meaningfulness. Figure 3 illustrates the evaluation process. Submission forms, clinical guidelines and supporting evidence, validity and reliability testing and any relevant eCQM materials were reviewed and evaluated based on the criteria outlined for meaningfulness in the PRMR Guidebook of Policies and Procedures.

Table 2.3.1 provides a detailed review of the evaluation criteria used by staff in developing the preliminary assessment.

Figure 3. Evaluation Process

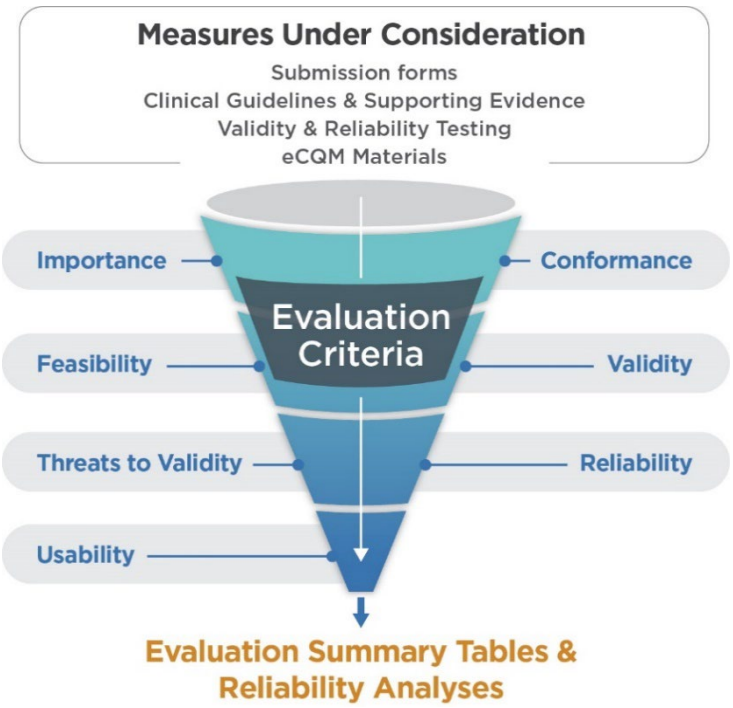


Table 2.3.1. Evaluation Criteria

Evaluation Criteria	Guiding Question
Concept of Interest	
Importance	*Does the measure align with interested party goals and priorities?*
Conformance	*Does the measure as specified align with the conceptual intent?*
Feasibility	*Does the measure specification and data collection minimize burden?*
Context of Use	
Importance	*Will performance improvement to the benchmark have a significant impact on population outcomes?*
Reliability	*Is measure performance scientifically sound?*
Validity	*May providers/facilities/care systems effectively improve on this measure?*
Threats to Validity	*If appropriate, is the measure risk adjusted to account for factors outside entity control?*
Usability	*Is there opportunity for improvement on this measure in the intended use setting?*

2.4 Data Analysis

Battelle staff reviewed and evaluated validity and reliability testing results provided in submission materials. Additionally, when reliability testing results were available, a team of analysts simulated median reliability to assess performance score deciles and reliability deciles and to generate mean reliability. The distribution of reliability across entities is important, and denominator size (generally patient population) has a great impact on reliability for a single entity. This information is not currently requested from the developer, but the data provided in the measure report and supplemental materials are used to simulate a dataset that closely mirrors any mean, standard deviation, and percentile information provided for the performance score or for reliability. Where possible, tables containing results of reliability analyses follow the measure evaluation tables for each measure. These values were generated through the following process and correspond to the order in which tables are shown:

1) Entities are sorted by performance score, and the average score by decile (estimated from the simulated data) is listed along with the number of entities and episodes included in each average. Average, standard deviation, and minimum and maximum scores are also included.

2) Entities are sorted by the number of episodes, and the average reliability by decile (estimated from the simulated data) is reported along with the number of entities and episodes included in each average and the average number of episodes per decile.

3) Entities are sorted by reliability, and the average reliability by decile (estimated from the simulated data) is reported. Average, standard deviation, and minimum and maximum reliability and inter-quartile range (IQR) are also included.

Battelle uses a reliability threshold of 0.6 for individual entities in these analyses, which aligns with reliability thresholds used across other CMS initiatives. In some instances, developers provided reliability-by-decile tables for inclusion in the report. These measures have footnotes to inform PRMR committee members if a table was derived via Battelle's simulated reliability analyses or was provided by the measure developer and derived from original testing data.

PRMR committee members should note that there is variation in the types of testing and data availability expected for measures at different stages of use and measure type. For example, when compared to in-use measures that are undergoing substantial changes, new measures do not have measure information forms and may have less robust testing and use available data. The history of each measure's endorsement pathway and inclusion in CMS programs is noted in the background section for each measure to guide PRMR committee members in their review. The appropriate testing methodology for validity and reliability may vary by measure type, and some measures may not be well-suited to utilizing risk adjustment models. Methods such as empiric validity were also not required as part of MUC submission but may provide stronger evidence of measure performance and suitability where submitted. When evaluators note that testing scores, clinical guidelines, or other information is absent from submitted materials, PRMR committee members should focus on the available information and direct their reviews toward possible implementation of each measure for the selected program.

Table 2.4.1 provides a summary of data sources that were submitted through CMS MERIT and reviewed, and the kinds of evidence and analyses presented in each submission. The focus in

the table is on testing performed at the measured-entity level, and the type of testing performed is noted.

Table 2.4.1. Data Sources for Clinician Committee Measures Under Consideration

MUC ID	Measure Title	Data Reviewed	Data Not Available
MUC2023-137	Initial Opioid Prescribing for Long Duration (IOP-LD)	✓ MERIT Submission Form ✓ Face Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature ✓ Clinical Practice Guidelines	× Measure Information Form × Empiric Validity
MUC2023-141	Positive PD-L1 Biomarker Expression Test Result Prior to First-Line Immune Checkpoint Inhibitor Therapy	✓ MERIT Submission Form ✓ Face Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature	× Measure Information Form × Clinical Practice Guidelines × Empiric Validity
MUC2023-161	Appropriate Germline Testing for Ovarian Cancer Patients	✓ Reliability: Signal-to-noise ✓ Peer-Reviewed Literature ✓ Empiric Validity: Pearson's Correlation Coefficient	× Measure Information Form × Clinical Practice Guidelines
MUC2023-162	Patient-Reported Pain Interference Following Chemotherapy among Adults with Breast Cancer	✓ MERIT Submission Form ✓ Empiric Validity: Pearson's Correlation Coefficient ✓ Face Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature ✓ Clinical Practice Guidelines	× Measure Information Form
MUC2023-164	Adult COVID-19 Vaccination Status	✓ MERIT Submission Form ✓ Face Validity	× Measure Information Form × Clinical Practice Guidelines × Empiric Validity

MUC ID	Measure Title	Data Reviewed	Data Not Available
		✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature	
MUC2023-179	Initiation and Engagement of Substance Use Disorder Treatment (IET)	✓ MERIT Submission Form ✓ Empiric Validity: Pearson's Correlation Test ✓ Face Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature ✓ Clinical Practice Guideline	× Measure Information Form
MUC2023-190	Patient-Reported Fatigue Following Chemotherapy among Adults with Breast Cancer	✓ MERIT Submission Form ✓ Measure Information Form ✓ Empiric Validity: Construct Validity ✓ Face Validity ✓ Reliability: Signal-to-Noice ✓ Peer-Reviewed Literature ✓ Clinical Practice Guideline	--
MUC2023-199	Connection to Community Service Provider	✓ Peer Reviewed Literature ✓ MERIT Submission Form ✓ Reliability: Signal-to-Noise ✓ Empiric Validity: Adjusted Odds Ratio ✓ Face Validity	× Measure Information Form × Clinical Guidelines
MUC2023-201	Cataract Removal with Intraocular Lens (IOL) Implantation	✓ MERIT Submission Form ✓ Empiric Validity: Construct Validity	× Measure Information Form

MUC ID	Measure Title	Data Reviewed	Data Not Available
		✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature ✓ Clinical Practice Guidelines	
MUC2023-203	Chronic Kidney Disease	✓ MERIT Submission Form ✓ Empiric Validity: Construct Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature ✓ Clinical Practice Guideline	× Measure Information Form
MUC2023-204	End-Stage Renal Disease	✓ MERIT Submission Form ✓ Empiric Validity: Construct Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature ✓ Clinical Practice Guideline	× Measure Information Form
MUC2023-205	Inpatient (IP) Percutaneous Coronary Intervention (PCI)	✓ MERIT Submission Form ✓ Empiric Validity: Construct Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature ✓ Clinical Practice Guideline	× Measure Information Form
MUC2023-206	Kidney Transplant Management	✓ MERIT Submission Form ✓ Empiric Validity: Construct Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature	× Measure Information Form

MUC ID	Measure Title	Data Reviewed	Data Not Available
		✓ Clinical Practice Guideline	
MUC2023-207	Prostate Cancer	✓ MERIT Submission Form ✓ Empiric Validity: Construct Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature ✓ Clinical Practice Guideline	× Measure Information Form
MUC2023-208	Respiratory Infection Hospitalization	✓ MERIT Submission Form ✓ Empiric Validity: Construct Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature ✓ Clinical Practice Guideline	× Measure Information Form
MUC2023-209	Rheumatoid Arthritis	✓ MERIT Submission Form ✓ Empiric Validity: Construct Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature ✓ Clinical Practice Guideline	× Measure Information Form
MUC2023-210	Resolution of At Least 1 Health-Related Social Need	✓ Peer Reviewed Literature ✓ MERIT Submission Form ✓ Clinical Guidelines ✓ Reliability: Signal-to-Noise ✓ Empiric Validity: Adjusted Odds Ratio ✓ Face Validity	× Measure Information Form

MUC ID	Measure Title	Data Reviewed	Data Not Available
MUC2023-211	Melanoma: Tracking and Evaluation of Recurrence	✓ MERIT Submission Form ✓ Face Validity ✓ Reliability: Random Split-Half Correlation ✓ Clinical Practice Guideline ✓ Peer-Reviewed Literature	× Measure Information Form
MUC2023-212	Level I Denials Upheld Rate Measure	✓ MERIT Submission Form ✓ Empiric Validity: Predictive Validity ✓ Reliability: Signal-to-Noise ✓ Peer-Reviewed Literature ✓ Clinical Practice Guideline	× Measure Information Form

Chapter 3. Measures by CMS Program

Merit-based Incentive Payment System-Cost Program (MIPS)

3.1 MUC2023-201 Cataract Removal with Intraocular Lens (IOL) Implantation

Description: The Cataract Removal with Intraocular Lens (IOL) Implantation episode-based cost measure evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients who undergo a procedure for cataract removal with IOL implantation.

Measure Type: Cost/Resource Use

Level of Analysis: Clinician: Group only

Data Source(s): Administrative Data (non-claims); Claims Data

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Cataract Removal with Intraocular (IOL) Implantation episode-based cost measure to the MIPS measure set. The measure assesses clinicians' risk-adjusted costs to Medicare for cataract removal with IOL implantation procedures. This measure is a revised version of the Routine Cataract Removal with Intraocular (IOL) Implantation measure that has been in the MIPS measure set since 2019. The substantive updates of the measure include expanding the measure scope to no longer exclude patients with certain ocular conditions—such as macular degeneration, glaucoma, and diabetic eye disease—and to include the costs of additional clinically related services, such as pre-operative testing, additional telehealth services, durable medical equipment (DME), emergency department (ED) visits for ocular complaints, and durable medical equipment to ensure a more comprehensive evaluation. CMS is considering this measure for use in MIPS because there are opportunities to improve patient outcomes and reduce the cost to Medicare for cataract removal procedures, which is a very common procedure for the Medicare population as more than 14 million surgeries were performed amongst Medicare beneficiaries between 2011 and 2019. It also addresses what would otherwise be a measurement gap area, given that ophthalmologists are the most frequently attributed specialty and do not have other cost measures that focus on the type of care they provide.

Table 3.1.1. MUC2023-201 Brief Summary of Measure Information

CMS MERIT Submission Information MUC2023- 201	Description
Measure name	Cataract Removal with Intraocular Lens (IOL) Implantation
MUC ID	MUC2023-201
Cascade priority	Affordability and Efficiency
Measure steward	Centers for Medicare & Medicaid Services
Measure developer	Acumen, LLC
Program submitted to	Merit-based Incentive Payment System- Cost
Committee assigned to	Clinician
Is this a new measure in this year's MUC list?	No
If not a new measure, then describe the history of this measure in prior MUC list inclusion	This measure was submitted as MUC17-235 to the Merit-based Incentive Payment System, 2017-2018 and was reviewed by the Clinician Workgroup leading to a supportive recommendation.
Is the measure currently used in a CMS program?	Yes
If previously used, please describe the history of the measure in CMS program	The Routine Cataract Removal with Intraocular Lens (IOL) Implantation episode-based measure is currently in use in the Merit-Based Incentive Payment System with the measure ID of COST_IOL_1 and is a previous iteration of the Cataract Removal with Intraocular Lens (IOL) Implantation measure.
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	Section 101(f) of Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)
CBE endorsement status	Not Endorsed; related measure Routine Cataract Removal with IOL Implantation measure was endorsed by a CBE in 2019 (CBE #3509) but revised measure not yet submitted
Path to endorsement	Unknown
Measure Specification Details	
Measure description	The Cataract Removal with Intraocular Lens (IOL) Implantation episode-based cost measure evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients who undergo a procedure for cataract removal with IOL implantation. This procedural measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during each cataract removal episode from 60 days prior to the clinical event that opens, or "triggers," the episode through 90 days after the trigger.

CMS MERIT Submission Information MUC2023- 201	Description
Data source	Administrative Data (non-claims); Claims Data
Level of analysis	Clinician: Group only
Numerator	The measure numerator is the ratio of the winsorized standardized observed cost to the expected cost for all Cataract Removal with Intraocular Lens (IOL) Implantation episodes attributed to a clinician. This sum is then multiplied by the national average winsorized observed episode cost to generate a dollar figure.
Denominator	The measure denominator is the total number of Cataract Removal with Intraocular Lens (IOL) Implantation episodes assigned to the clinician.
Numerator exclusions	N/A
Denominator exclusions	The following standard exclusions are applied to ensure data completeness: • Patient has a primary payer other than Medicare for any time overlapping the episode window or 120-day lookback period prior to the trigger day. • Patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window or was enrolled in Part C for any part of the lookback plus episode window. • No main clinician is attributed the episode. • Patient's date of birth is missing. • Patient's death date occurred before the episode ended. • The episode trigger claim was not performed in an OP hospital or ASC setting based on its place of service. Exclusions specific to the Cataract Removal with Intraocular Lens (IOL) Implantation measure are developed with input from the Cataract Removal with Intraocular Lens (IOL) Implantation Clinician Expert Workgroup and include patients with significant ocular conditions impacting surgical complication rate/visual outcomes.
Denominator exceptions	N/A
Risk adjustment	Yes
Development Status	Fully Developed
Target population	Medicare Fee-for-Service (FFS)
Measure type	Cost/Resource Use
Is the measure a composite or component of a composite?	No

CMS MERIT Submission Information MUC2023 201	Description
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.1.2. MUC2023-201 Measure Evaluation

MUC2023 201 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	Among Medicare beneficiaries, 14,396,438 cataract surgeries were performed between 2011 and 2019.[4] IOL procedures have a relatively low rate of complications. However, complications that require long-term management or result in a return to the operating room persist.[5]	The measure developer did not submit empirical evidence linking complications to episode-based costs.	The study population is the same as the target quality program population.

[4] Zafar, S., Dun, C., Srikumaran, D., Wang, P., Schein, O. D., Makary, M., & Woreta, F. (2022). Endophthalmitis Rates among Medicare Beneficiaries Undergoing Cataract Surgery between 2011 and 2019. Ophthalmology, 129(3), 250-257. https://doi.org/10.1016/j.ophtha.2021.09.004

[5] Schmier, J. K., Hulme-Lowe, C. K., Covert, D. W., & Lau, E. C. (2016). An updated estimate of costs of endophthalmitis following cataract surgery among Medicare patients: 2010-2014. Clin Ophthalmol, 10, 2121-2127. https://doi.org/10.2147/opth.S117958

MUC2023 201 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	The American Academy of Ophthalmology (AAO) preferred practice pattern recommends surgeons be aware of and prepared to manage high-risk characteristics that may complicate cataract surgery.[6] Variables identified in the literature search related to complications include materials used in the procedure, IOL's being placed asymmetrically within a patient's eye, and patient-level characteristics such as age. Furthermore, some antibiotics are associated with a greater risk reduction of endophthalmitis than others.[7] Unnecessary preoperative testing can result in high IOL procedure-related costs. Preoperative testing has been found ineffective in reducing the risk of adverse events or improving outcomes.[8] However, an observational study of Medicare beneficiaries undergoing cataract surgery in 2011 found that 53% had at least one		

[6] American Academy of Ophthalmology. Cataract in the Adult Eye Preferred Practice Pattern. American Academy of Ophthalmology; Chicago, IL, USA: 2016. Available online: https://gmcboard.vermont.gov/sites/gmcb/files/Cataract%20in%20the%20Adult%20Eye%20PPP.pdf
[7] Moser, C. L., Lecumberri Lopez, M., Garat, M., & Martín-Baranera, M. (2019). Prophylactic intracameral cefazolin and postoperative topical moxifloxacin after cataract surgery: endophthalmitis risk reduction and safety results in a 16-year study. Graefes Arch Clin Exp Ophthalmol, 257(10), 2185-2191. https://doi.org/10.1007/s00417-019-04417-9
[8] American Academy of Ophthalmology. Cataract in the Adult Eye Preferred Practice Pattern. American Academy of Ophthalmology; Chicago, IL, USA: 2016. Available online: https://gmcboard.vermont.gov/sites/gmcb/files/Cataract%20in%20the%20Adult%20Eye%20PPP.pdf

MUC2023 201 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	preoperative test the month before surgery, resulting in higher expenditures on testing and office visits, at $4.8 million and $12.4 million, respectively.[9]		
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	The Cataract Removal with IOL Implantation cost measure can be triggered based on claims data from: ambulatory surgical centers (ASC) and hospital outpatient departments (HOPD). The measure has exclusions to ensure data completeness.	No empirical evidence of conformance was submitted. This measure excludes episodes for patients with ocular conditions that could impact visual outcomes or surgical complication rates. However, empirical analyses by the measure developer shows many of the excluded episodes are similar to those included in the measure. As part of the comprehensive reevaluation, certain exclusion criteria were removed from the revised Cataract Removal with IOL implantation measure in alignment with the MMS Blueprint guidance.	Some persons and entities in the quality program population are included in the specification. Some data element reliability and validity analyses extrapolate to the quality program population.
Feasibility: Does the measure's specification and data	This is a claims-based measure. It can be calculated from claims submitted electronically for billing and information collected for other purposes. All data		The people, processes, and technology required for data collection and reporting extrapolate to the quality program population.

[9] Schmier, J. K., Hulme-Lowe, C. K., Covert, D. W., & Lau, E. C. (2016). An updated estimate of costs of endophthalmitis following cataract surgery among Medicare patients: 2010-2014. Clin Ophthalmol, 10, 2121-2127. https://doi.org/10.2147/opth.S117958

MUC2023 201 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
collection minimize burden? *(Concept of Interest)*	elements are in defined fields in electronic sources.		Some entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Will performance improvement to the benchmark have a significant impact on population outcomes? (*Context of Use)*	Data submitted by the measure developer show variability in measure performance. For TINs that meet a 10 episode-case minimum, the minimum performance score is $1,227.35, the 10th percentile score is $2,858.02, overall mean performance score is $3,171.92 (standard deviation = $394.97), the median is $3,092.73, the 90th percentile score is $3,622.81, and the maximum score is $5,070.42.	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound? *(Context of Use)*	According to the data submitted by the measure developer, at the 10-episode testing volume threshold, the mean reliability for the Cataract Removal with IOL Implantation measure is high, precisely 0.97 and 0.96 at the TIN and TIN-NPI levels, respectively. Mean reliability levels above 0.7 demonstrate high reliability for cost measures as previously established in the CY 2017 Quality Payment Program final rule (81 FR 77169 through 77171).	--	Most or all entities have reliability above the threshold (0.60) *within* the quality program population.
Validity: May providers/facilities/care systems effectively improve on this measure?	Data submitted by the measure developer show the costs of outpatient evaluation and management, major or minor procedures, anesthesia, and other Part B-covered drugs are	The measure developer did not specify how an entity may improve performance on areas associated with high episode-based costs identified in their gap	There is an association between the entity and the measure focus *within* the quality program population.

MUC2023 201 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
(Context of Use)	associated with a worse measure score even after controlling for adverse events. This pattern suggests that, after controlling for the cost of adverse outcomes, these services remain cost drivers of cataract removal episodes. With the exception of minor procedures at the TIN-NPI reporting level, none of these services are associated with the cost of adverse events, which suggests that fewer of these services can be provided without increasing costs associated with adverse events.	analysis. The areas identified as associated with costs are (1) mitigation of costly complications that require long-term management, (2) mitigation of complications that result in a return to the operating room, and (3) reducing preoperative testing.	
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	The risk adjustment model for this measure uses a linear regression model, which utilizes variables from the CMS Hierarchical Condition Code Version 24 (CMS-HCC V24) 2021 Risk Adjustment Model. This includes comorbidities captured by 86 HCC codes that map with thousands of ICD-10-CM codes, interaction variables accounting for a range of comorbidities, and other standard risk adjustors, including patient level demographics such as age and health status (i.e., disability status, end-stage renal disease [ESRD] status, recent use of long-term care). Additional risk adjustors that are clinically relevant to this measure were developed with input from the Cataract	--	N/A

MUC2023 201 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	Removal with Intraocular Lens (IOL) Implantation Clinician Expert Workgroup. The measure is further stratified into sub-groups (i.e., bilateral cataract surgery in an ambulatory surgical center [ASC], unilateral cataract surgery in an ASC, bilateral cataract surgery in a hospital outpatient department [HOPD], and unilateral cataract surgery in an HOPD). Risk adjustment is performed separately for episodes within each sub-group of this measure to allow for comparisons within more clinically homogenous cohorts.		
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	The American Academy for Ophthalmology (AAO) preferred practice pattern provides best practice guidelines that should reduce complication rates (which should be related to episode-based costs, but no empirical evidence of this association was included in submitted materials) and pre-operative testing (which is related to preoperative costs).[10]	The submitted materials do not specify the resources and context that might facilitate or be a barrier to the way an entity may improve.	Unable to determine if there is an articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

[10] American Academy of Ophthalmology. Cataract in the Adult Eye Preferred Practice Pattern. American Academy of Ophthalmology; Chicago, IL, USA: 2016. Available online: https://gmcboard.vermont.gov/sites/gmcb/files/Cataract%20in%20the%20Adult%20Eye%20PPP.pdf

MUC2023-201 Measure Reliability

The performance score is a risk- and specialty-adjusted average cost per episode.

For TIN-level entities with at least 10 episodes, the information provided indicates a median signal-to-noise reliability of 0.97across 4,080 entities.

For TIN-level entities with at least 20 episodes, the information provided indicates a median signal-to-noise reliability of 0.98 across 3,704 entities.

For TIN-level entities with at least 30 episodes, the information provided indicates a median signal-to-noise reliability of 0.98 across 3,329 entities.

Decile tables:

The developer provided decile tables of performance score for the 10-episode threshold level and decile tables for reliability for the 10-, 20-, and 30-episode threshold levels.

For Tables 3.1.3, entities are sorted by performance score, and the average score by decile is reported along with the number of entities included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.1.3. MUC2023-201 Performance Score Deciles – TINs with at least 10 attributed episodes

MUC2023 210	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score (SD)	$3,172 (394)	1,227	2,858	2,944	2,993	3,038	2,093	3,175	3,254	3,373	3,623	4,608	5,070
Entities	4,080	1	408	408	408	408	408	408	408	408	408	408	1
# Episodes	800,983	--	--	--	--	--	--	--	--	--	--	--	--

For Tables 3.1.4, 3.1.5 and 3.1.6, entities were sorted by reliability, and the average reliability by decile is reported. Average and standard deviation are also included.

Table 3.1.4. MUC2023-201 Mean Reliability (by Reliability Decile) – TINs with at least 10 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.97	0.034	0.914	0.945	0.962	0.972	0.980	0.985	0.989	0.993	0.996	0.999

Table 3.1.5. MUC2023-201 Mean Reliability (by Reliability Decile) – TINs with at least 20 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.98	0.021	0.939	0.958	0.969	0.977	0.982	0.987	0.990	0.993	0.996	0.999

Table 3.1.6. MUC2023-201 Mean Reliability (by Reliability Decile) – TINs with at least 30 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.98	0.015	0.956	0.967	0.975	0.980	0.985	0.988	0.991	0.994	0.996	0.999

Interpretation:

With a threshold of at least 10 episodes the median reliability is 0.97. The mean reliability of the lowest decile is 0.914, so very few, if any, of the entities have reliability below 0.6.

3.2 MUC2023-203 Chronic Kidney Disease

Description: The Chronic Kidney Disease (CKD) 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 stage 4 or 5 chronic kidney disease.

Measure Type: Cost/Resource Use

Level of Analysis: Clinician Group

Data Source(s): Administrative Data (non-claims); Claims Data

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the chronic kidney disease (CKD) episode-based cost measure to the MIPS measure set. The measure assesses 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 chronic kidney disease stages 4 and 5 (not on dialysis). Chronic kidney disease is a priority area for CMS and accounts for high expenditures. This measure, in conjunction with the End-Stage Renal Disease and Kidney Transplant Management Disease measures that CMS is also considering, will fill a measurement gap area in the care continuum for a patient with kidney disease as there are currently no existing cost measures for this clinical area. CMS is considering this measure for use in MIPS because it has the potential to capture many patients and Medicare spending and because there are opportunities for clinicians to improve their performance, such as reducing downstream costs of hospitalizations and readmissions through care coordination.

Table 3.2.1. MUC2023-203 Measure Information

MERIT Submission Information MUC2023-203	Description
Measure name	Chronic Kidney Disease
MUC ID	MUC2023-203
Cascade priority	Affordability and Efficiency
Measure steward	Centers for Medicare & Medicaid Services

MERIT Submission Information MUC2023-203	Description
Measure developer	Acumen, LLC
Program submitted to	Merit-based Incentive Payment System-Cost
Committee assigned to	Clinician Committee
Related measures in the program	N/A
Is this a new measure in this year's MUC List?	Yes
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program?	No
If previously used, please describe the history of the measure in CMS program	N/A
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	Section 101(f) of Medicare Access and CHIP Reauthorization Act of 2015 (MACRA).
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
Measure specification details	
Measure description	The Chronic Kidney Disease (CKD) 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 stage 4 or 5 chronic kidney disease. This chronic condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during a chronic kidney disease episode.
Data source	Administrative Data (non-claims); Claims Data
Level of analysis	Clinician: Group only
Numerator	The measure numerator is the weighted average ratio of the winsorized scaled standardized observed cost to the scaled expected cost for all chronic kidney disease episodes attributed to a clinician, where each ratio is weighted by each

MERIT Submission Information MUC2023-203	Description
	episode's number of days assigned to a clinician. This sum is then multiplied by the national average winsorized scaled observed episode cost to generate a dollar figure.
Denominator	The measure denominator is the total number of days from chronic kidney disease episodes assigned to the clinician across all patients.
Numerator exclusions	N/A
Denominator exclusions	The following standard exclusions are applied to ensure data completeness: • Patient has a primary payer other than Medicare for any time overlapping the episode window or 365-day lookback period prior to the episode window. • Patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window or was enrolled in Part C for any part of the lookback plus episode window. • Patient was not found in the Medicare Enrollment Database (EDB). • Patient's death date occurred before the episode end date. • Patient has an episode window shorter than 1 year. • Patients with extremely low treatment costs. • Patient's residence is outside the United States or its territories during the episode window, as indicated in the EDB. • There are no exclusions specific to the chronic kidney disease measure.
Denominator exceptions	N/A
Risk adjustment	Yes
Development stage	Fully Developed
Target population	Medicare Fee-for-Service (FFS)
Measure type	Cost/Resource Use
Is the measure a composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No

MERIT Submission Information MUC2023 203	Description
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.2.2. MUC2023-203 Measure Evaluation

MUC2023 203 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Importance: Does the measure align with goals and priorities? *(Concept of Interest)*	Studies cited show relevance of cost containment related to stage 4 and 5 CKD and progression to End-Stage Renal Disease (ESRD).[11,12] Implementing a value-based approach to CKD and kidney care more broadly may help incentivize clinicians better to manage patients' CKD progression and transition to ESRD. Gap analysis submitted by the developer identified evidence of value-based care at the clinician level being effective in improving patient care through the ESRD Seamless Care Organization program.[13]	No evidence submitted directly showing mechanisms by which clinicians will improve CKD care and slow progression to ESRD through implementation of cost containment measures.	The study population is the same as the target quality program population.

[11] Liu HH, Zhao S. Savings Opportunity from Improved CKD Care Management. *J Am Soc Nephrol*. 2018;29(11):2612-2615. doi:10.1681/ASN.2017121276.
[12] Spencer D, Dunning S, McPheeters J, St Clair Russell J, Hane C. Health care costs associated with unrecognized progression to late-stage kidney disease. *Am J Manag Care*. 2023;29(2):e64-e68. Published 2023 Feb 1. doi:10.37765/ajmc.2023.89323.
[13] Johnson DS, Meyer KB. Leading Integrated Kidney Care Entities of the Future. *Adv Chronic Kidney Dis*. 2018;25(6):523-529. doi:10.1053/j.ackd.2018.09.001.

MUC2023 203 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	This measure's specifications align with the intent of the measure. It has appropriate attribution at clinician group level and includes in specification, which includes observed cost to the scaled expected cost for all chronic kidney disease episodes attributed to a clinician, where each ratio is weighted by each episode's number of days assigned to a clinician.	--	Most persons and entities in the quality program population are included in the specification. Data element reliability and validity extrapolate to the quality program population.
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	This is a claims-based measure. It can be calculated exclusively from claims submitted electronically for billing and other purposes. All data elements are in defined fields in electronic sources.	--	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Will performance improvement to the benchmark have a significant impact on population outcomes? *(Context of Use)*	Evidence provided demonstrated that performance measurement aimed at assessing provider accountability may help to drive improvements on the measure target. As specified, improvements on the measure benchmark will likely have significant impact on population outcomes such as episode-based cost but also potentially quality of life.	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound? *(Context of Use)*	This measure underwent reliability testing to evaluate the measure's ability to consistently differentiate one clinician's performance from another.	With a threshold of at least 10 episodes per entity, the median reliability is 0.293. The reliability for about 75% of the entities is below 0.6. When the threshold is increased to a minimum of 20 episodes per	Some entities have reliability above the threshold (0.60) *within* the quality program population OR a population that can be extrapolated to the program population.

MUC2023 203 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
		entity, the median reliability is 0.386. The reliability for about 75% of the entities is still below 0.6. When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.448. About 65% of the entities may have reliability below 0.6.	
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	Validity was evaluated empirically by estimating the effect of relevant treatment choices on the measure score. There was a strong association between measure focus and measured entity (clinician group). Clinician groups can assert reasonable influence and improve outcomes on this measure based on associations seen in empiric testing.	While there was not a clear plan for how an entity may improve performance on the measure focus, the treatment choice assessments during validity testing suggest that such a plan could easily be developed from available data.	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population. There is limited articulation of the way an entity may improve performance on the measure focus within the quality program population.
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	This measure is risk adjusted and there is an explicit rationale for confounders included in the model. Risk adjustment uses CMS Hierarchical Condition Code Version 24 (CMS-HCC V24) 2021 Risk Adjustment Model. This includes comorbidities captured by 86 HCC codes that map with thousands of ICD-10-CM codes and other standard risk adjustors, including interaction variables accounting for a range of comorbidities, patient level demographics such as age and health	--	N/A

MUC2023 203 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	status (i.e., disability status, end-stage renal disease [ESRD] status, recent use of long-term care), HCC count, patient dual eligibility status, and number and types of clinician specialties from which the patient has received care.		
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	This measure has the opportunity for improvement at the clinician group level. The measure's development is aligned with episode-based cost measures currently used in the program. The CKD measure was developed in consideration of alignment opportunities with CMS' Kidney Care First (KCF) and Comprehensive Kidney Care Contracting (CKCC) payment Options of the KCC Advanced Payment Model, which will improve usability.	There is not an explicit articulation of the resources and context that might facilitate or be a barrier to the way an entity may improve.	There is an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-203 Measure Reliability

The performance score is a risk- and specialty-adjusted average cost per episode.

Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across all entities. σ^2_{within} is the standard deviation of the score for a single entity.

For TIN-level entities with at least 10 episodes, the information provided indicates a median signal-to-noise reliability of 0.293 across 3,698 entities.

For TIN-level entities with at least 20 episodes, the information provided indicates a median signal-to-noise reliability of 0.386 across 2,301 entities.

For TIN-level entities with at least 30 episodes, the information provided indicates a median signal-to-noise reliability of 0.448 across 1,743 entities.

Decile tables:
The developer provided decile tables of performance score for the 20-episode threshold level and decile tables for reliability for the 10-, 20-, and 30-episode threshold levels.

For Table 3.2.3, entities were sorted by performance score, and the average score by decile is reported along with the number of entities included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.2.3. MUC2023-203 Performance Score Deciles – TINs with at least 20 attributed episodes

MUC2023-203	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	8,654 (1,966)	2,905	6,360	7,106	7,620	8,087	8,489	8,919	9,402	9,967	11,068	14,782	21,015
Entities	2,301	1	230	230	230	230	230	230	230	230	230	230	1
# Episodes	262,192	--	--	--	--	--	--	--	--	--	--	--	--

For Tables 3.2.4, 3.2.5, and 3.2.6, entities are sorted by reliability and the average reliability by decile is reported. Average and standard deviation are also included.

Table 3.2.4. MUC2023-203 Mean Reliability (by Reliability Decile) – TINs with at least 10 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.329	0.209	0.081	0.130	0.179	0.231	0.293	0.357	0.435	0.528	0.636	0.837

Table 3.2.5. MUC2023-203 Mean Reliability (by Reliability Decile) – TINs with at least 20 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.405	0.196	0.161	0.218	0.274	0.329	0.386	0.448	0.518	0.589	0.683	0.848

Table 3.2.6. MUC2023-203 Mean Reliability (by Reliability Decile) – TINs with at least 30 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.455	0.184	0.216	0.283	0.343	0.392	0.448	0.504	0.555	0.622	0.711	0.859

Interpretation:

With a threshold of at least 10 episodes per entity, the median reliability is 0.293. About 75% of the entities may have reliability below 0.6.

When the threshold is increased to a minimum of 20 episodes per entity, the median reliability is 0.386. About 75% of the entities may have reliability below 0.6.

When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.448. About 65% of the entities may have reliability below 0.6.

3.3 MUC2023-204 End-Stage Renal Disease

Description: The End-Stage Renal Disease (ESRD) 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 ESRD.

Measure Type: Cost/Resource Use

Level of Analysis: Clinician Group

Data Source(s): Administrative Data (non-claims); Claims Data

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the End-Stage Renal Disease (ESRD) episode-based cost measure to the MIPS measure set. The measure assesses a clinician's or clinician group's risk adjusted and specialty-adjusted costs to Medicare for patients who receive medical care to manage ESRD. Patients with ESRD receiving dialysis care are a vulnerable population, and dialysis care accounts for significant expenditures over a potentially long period of time. The ESRD measure, in conjunction with the Chronic Kidney Disease and Kidney Transplant Management Disease measures that CMS is also considering, will fill a measurement gap area in the care continuum for a patient with kidney disease as there are currently no existing cost measures for this clinical area. CMS is considering this measure for use in MIPS because it has the potential to capture a large number of patients and Medicare spending and because there are opportunities for clinicians to improve their performance, such as reducing hospitalizations and using appropriate dialysis care.

Table 3.3.1. MUC2023-204 Brief Summary of Measure Information

MERIT Submission Information MUC2023-204	Description
Measure name	End-Stage Renal Disease
MUC ID	MUC2023-204
Cascade priority	Affordability and Efficiency
Measure steward	Centers for Medicare & Medicaid Services
Measure developer	Acumen, LLC
Program submitted to	Merit-based Incentive Payment System-Cost

MERIT Submission Information MUC2023-204	Description
Committee assigned to	Clinician Committee
Related measures in the program	N/A
Is this a new measure in this year's MUC List?	Yes
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program?	N/A
If previously used, please describe the history of the measure in CMS program	New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program.
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	Section 101(f) of Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
Measure Specification Details	
Measure description	The End-Stage Renal Disease (ESRD) 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 ESRD. This chronic condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during an ESRD episode.
Data source	Administrative Data (non-claims); Claims Data
Level of analysis	Clinician: Group only
Numerator	The measure numerator is the weighted average ratio of the winsorized scaled standardized observed cost to the scaled expected cost for all End-Stage Renal Disease episodes attributed to a clinician, where each ratio is weighted by each episode's number of days assigned to a clinician. This sum is then multiplied by the national average winsorized scaled observed episode cost to generate a dollar figure.
Denominator	The measure denominator is the total number of days from End-Stage Renal Disease episodes assigned to the clinician across all patients.
Numerator exclusions	N/A
Denominator exclusions	The following standard exclusions are applied to ensure data completeness:

MERIT Submission Information MUC2023-204	Description
	• Patient has a primary payer other than Medicare for any time overlapping the episode window or 365-day lookback period prior to the episode window. • Patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window or was enrolled in Part C for any part of the lookback plus episode window. • Patient was not found in the Medicare Enrollment Database. • Patient's death date occurred before the episode end date. • Patient has an episode window shorter than 1 year. • Patients with extremely low treatment costs. • Patient's residence is outside the United States or its territories during the episode window, as indicated in the Enrollment Database. Exclusions specific to the End-Stage Renal Disease measure are developed with input from the Chronic Kidney Disease/End-Stage Renal Disease Clinician Expert Workgroup and include episodes that ended in kidney transplant.
Denominator exceptions	N/A
Risk adjustment	Yes
Development status	Fully Developed
Target population	Medicare Fee-for-Service (FFS)
Measure type	Cost/Resource Use
Is the measure a composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.3.2. MUC2023-204 Measure Evaluation

MUC2023-204 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	Studies cited show relevance of cost containment related to management of End-Stage Renal Disease (ESRD).[14,15,16] Implementing a value-based approach to ESRD management more broadly may help incentivize clinicians better to manage quality of life and adoption of appropriate dialysis care for ESRD patients. Submitted initial findings from Medicare's Comprehensive End-Stage Renal Disease Care (CEC) model suggested that value-based purchasing programs can influence care coordination, and is associated with improved patient outcomes, reducing downstream costs.[17,18]	No evidence submitted directly showing mechanisms by which clinicians will improve CKD care and slow progression to ESRD through implementation of cost containment measures.	The study population is the same as the target quality program population.

[14] United States Renal Data System. *2022 USRDS Annual Data Report: Epidemiology of kidney disease in the United States*.
[15] Liu HH, Zhao S. Savings Opportunity from Improved CKD Care Management. *J Am Soc Nephrol*. 2018;29(11):2612-2615. doi:10.1681/ASN.2017121276.
[16] Spencer D, Dunning S, McPheeters J, St Clair Russell J, Hane C. Health care costs associated with unrecognized progression to late-stage kidney disease. *Am J Manag Care*. 2023;29(2):e64-e68. Published 2023 Feb 1. doi:10.37765/ajmc.2023.89323.
[17] Ullman, Darin F., Gregory J. Boyer, Brighita Negrusa, Richard A. Hirth, Jennifer Wiens, and Grecia Marrufo. "Medicare's Specialty-Oriented Accountable Care Organization: First-Year Results For People With End-Stage Renal Disease: Study examines Medicare's specialty-oriented accountable care organization program for people with end-stage renal disease." *Health Affairs* 41, no. 6 (2022): 893-900.
[18] Hirth, Richard A., Tammie Nahra, Jonathan H. Segal, Joseph Gunden, Grecia Marrufo, Brighita Negrusa, Gregory Boyer et al. "Association of the Comprehensive ESRD Care Model with Treatment Adherence." *Kidney360* 3, no. 6 (2022): 1039-1046.

MUC2023-204 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	This measure's specifications align with the measure's intent. It has appropriate attribution at clinician group level and includes in specification which includes observed cost to the scaled expected cost for all ESRD management episodes attributed to a clinician, where each ratio is weighted by each episode's number of days assigned to a clinician.	--	Most persons and entities in the quality program population are included in the specification. Data element reliability and validity extrapolate to the quality program population.
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	This is a claims-based measure. It can be calculated exclusively from claims submitted electronically for billing and other purposes. All data elements are in defined fields in electronic sources.	No explicit articulation of people, processes, or technology required.	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Will performance improvement to the benchmark have a significant impact on population outcomes? (*Context of Use)*	Evidence provided demonstrated that performance measurement aimed at assessing provider accountability may help to drive improvements on the measure target. As specified, improvements on the measure benchmark will likely have significant impact on population outcomes such as episode-based cost but also potentially quality of life.	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound? *(Context of Use)*	Appropriate signal-to-noise reliability testing was conducted in a relevant study population.	With a threshold of at least 10 episodes per entity, the median reliability is 0.403. The reliability for about 75% of the entities is below 0.6. When the threshold is increased to a minimum of 20 episodes per	Some entities have reliability above the threshold (0.60) *within* the quality program population OR a population that can be extrapolated to the program population.

MUC2023-204 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
		entity, the median reliability is 0.511. The reliability for about 65% of the entities is below 0.6. When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.571. Over 50% of the entities have a reliability below 0.6.	
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	Validity was evaluated empirically by estimating the effect of relevant treatment choices on the measure score. There was a strong association between measure focus and measured entity (clinician group). Clinician groups can assert reasonable influence and improve outcomes on this measure based on associations seen in empiric testing.	While there was not a clear plan for the explicit articulation of the way an entity may improve performance on the measure focus, the treatment choice assessments during validity testing suggest that such a plan could easily be developed from data available.	There is an association between the entity and the measure focus *within* the quality program population.
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	This measure is risk adjusted and there is an explicit rationale for confounders included in the model. Risk adjustment uses CMS Hierarchical Condition Code Version 24 (CMS-HCC V24) 2021 Risk Adjustment Model. This includes comorbidities captured by 86 HCC codes that map with thousands of ICD-10-CM codes and other standard risk adjustors, including interaction variables accounting for a range of comorbidities, patient level demographics (i.e., age) and health status (i.e., disability status, end-stage renal disease [ESRD] status, recent	--	N/A

MUC2023-204 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	use of long-term care), HCC count, patient dual eligibility status, and number and types of clinician specialties from which the patient has received care.		
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	This measure has the opportunity for improvement at the clinician group level. The measure's development is aligned with episode-based cost measures currently used in the program. The ESRD measure was also developed in consideration of alignment opportunities with CMS' KCF and CKCC payment Options of the KCC Advanced Payment Model.	There is not an explicit articulation of the resources and context that might facilitate or be a barrier to the way an entity may improve.	There is an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-204 Measure Reliability[19]

The performance score is a risk- and specialty-adjusted average cost per episode.

Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across all entities. σ^2_{within} is the standard deviation of the score for a single entity.

For TIN-level entities with at least 10 episodes, the information provided indicates a median signal-to-noise reliability of 0.403 across 3,142 entities.

For TIN-level entities with at least 20 episodes, the information provided indicates a median signal-to-noise reliability of 0.511 across 2,041 entities.

[19] Developer provided decile values to replace simulated reliability tables.

For TIN-level entities with at least 30 episodes, the information provided indicates a median signal-to-noise reliability of 0.571 across 1,516 entities.

Decile tables:

The developer provided decile tables of performance score for the 20-episode threshold level and decile tables for reliability for the 10-, 20-, and 30-episode threshold levels.

For Table 3.3.3, entities are sorted by performance score, and the average score by decile is reported along with the approximate number of entities included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.3.3. MUC2023-204 Performance Score Deciles – TINs with at least 20 attributed episodes

MUC2023 204	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	63,578 (4,685)	49,912	58,130	59,964	61,224	62,251	63,282	64,250	65,488	66,795	69,479	77,355	99,537
Entities	2,041	1	204	204	204	204	204	204	204	204	204	204	1
# Episodes	204,226	--	--	--	--	--	--	--	--	--	--	--	--

For Tables 3.3.4,3,3,5, and 3.3.6, entities are sorted by reliability, and the average reliability by decile is reported. Average and standard deviation are also included.

Table 3.3.4. MUC2023-204 Mean Reliability (by Reliability Decile) – TINs with at least 10 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.425	0.211	0.160	0.222	0.282	0.341	0.403	0.472	0.543	0.630	0.725	0.886

Table 3.3.5. MUC2023-204 Mean Reliability (by Reliability Decile) – TINs with at least 20 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.518	0.186	0.274	0.343	0.399	0.457	0.511	0.564	0.630	0.691	0.775	0.908

Table 3.3.6. MUC2023-204 Mean Reliability (by Reliability Decile) – TINs with at least 30 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.577	0.167	0.361	0.426	0.477	0.523	0.571	0.628	0.679	0.730	0.798	0.922

Interpretation:

With a threshold of at least 10 episodes per entity, the median reliability is 0.403. About 75% of the entities may have a reliability below 0.6.

When the threshold is increased to a minimum of 20 episodes per entity, the median reliability is 0.511. About 65% of the entities may have a reliability below 0.6.

When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.571. Over 50% of the entities have a reliability below 0.6.

These results suggest this measure has a low capability of distinguishing the quality of performance between entities.

3.4 MUC2023-190 Patient-Reported Fatigue Following Chemotherapy among Adults with Breast Cancer

Description: The PRO-PM will assess fatigue following chemotherapy administered with curative intent to adult patients with breast cancer.

Measure Type: PRO-PM or Patient Experience of Care

Level of Analysis: Clinician: Individual and Group

Data Source(s): Electronic Health Record; Paper Medical Records; PROMIS Scale: Fatigue Short Form 4a; Patient-Reported Data and Surveys

Development Status: Fully Developed

Endorsement Status: Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Patient-Reported Fatigue Following Chemotherapy among Adults with Breast Cancer measure to the MIPS quality measure set. This patient-reported outcome-based performance quality measure fills a current CMS high priority measure inventory gap within the oncologic clinical topic and is a concept not duplicative of quality measures currently in MIPS. Additionally, this measure addresses the patient experience of care for those with breast cancer with fatigue experienced following chemotherapy. Data from this measure provides insight into the effectiveness of medical oncologists in helping patients to minimize the persistent impact of their treatments and is useful to inform practice improvement. While the Oncology specialty measure set includes 25 measures, only 11 are specialty specific. This would be the first outcome specialty specific oncology measure to address the patient experience of care within this measure set. There is potential consideration for adding broader cancer diagnosis such as colon and lung cancer to this measure in the future.

Table 3.4.1. MUC2023-190 Brief Summary of Patient-Reported Fatigue Following Chemotherapy among Adults with Breast Cancer

MERIT Submission Information MUC2023-190	Description
Measure name	Patient-Reported Fatigue Following Chemotherapy among Adults with Breast Cancer
MUC ID	MUC2023-190
Cascade priority	Person-Centered Care
Measure steward	Purchaser Business Group on Health
Measure developer	Purchaser Business Group on Health

MERIT Submission Information MUC2023-190	Description
Program submitted to	Merit-based Incentive Payment System-Quality
Committee assigned to	Clinician Committee
Related measures in the program	N/A
Is this a new measure in this year's MUC list?	No
If not a new measure, then describe the history of this measure in prior MUC list inclusion	Submitted previously but not included in MUC List
Is the measure currently used in a CMS program	N/A
If previously used, please describe the history of the measure in CMS program	N/A
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	N/A
CBE endorsement status	Endorsed
CBE endorsement number if applicable	CBE 3720
History of endorsement	Year of most recent CDP endorsement: 2023
Path to endorsement	Year of next anticipated CDP endorsement review: 2026
Measure Specification Details	
Measure description	The PRO-PM will assess fatigue following chemotherapy administered with curative intent to adult patients with breast cancer.
Data source	Electronic Health Record; Paper Medical Records; PROMIS Scale: Fatigue Short Form 4a; Patient Reported Data and Surveys
Level of analysis	Clinician: Individual and Group (MIPS-Quality only)
Numerator	The PRO-PM numerator is the mean of the patient-level PROMIS Fatigue scores at the follow-up survey.
Denominator	Adult patients with stages I-III female breast cancer receiving an initial chemotherapy regimen. Denominator details: • The denominator population includes the following patients: • ≥ age 18 on the date of diagnosis, AND • Stages I-III female breast cancer AND • Receiving an initial chemotherapy regimen with a defined duration at the test site

MERIT Submission Information MUC2023-190	Description
	• Patients with baseline and follow-up PROMIS surveys (See the MIF/Data Dictionary for additional definitions)
Numerator exclusions	N/A
Denominator exclusions	• Patients on a therapeutic clinical trial. • Patients with recurrence/disease progression. • Patients who leave the practice during the follow-up period. • Patients who die during the follow-up period.
Denominator exceptions	N/A
Risk adjustment	Yes
Development Status	Fully Developed
If not fully developed, development stage	N/A
Target population	Medicare Fee for Service, Medicare Advantage, Medicaid, All Payer. All adult cancer patients not restricted by payer type.
Measure type	PRO-PM or Patient Experience of Care
Is the measure composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.4.2 Patient-Reported Fatigue Following Chemotherapy among Adults with Breast Cancer

MUC2023-190 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	Evidence suggests that fatigue is not only common among those undergoing cancer treatments, it has also been documented as the most distressing symptom following care.[20] Further, prevalence of fatigue is high, with 80% of patients receiving chemotherapy reporting experiencing fatigue.[21] Researchers have evaluated PROMIS pain and fatigue measures in cancer populations and evidence supports alignment of the measure focus with clinical goals and priorities.[22]	--	The study population is the same as the target quality program population.
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	Measure specification aligns with intent of use and is appropriate for the selected CMS program.	--	Most persons and entities in the quality program population are included in the specification. Data element reliability and validity extrapolate to the quality program population.

[20] Hinds, P.S., Quargnenti, A., Bush, A.J., Pratt, C., Fairclough, D., Rissmiller, G., Betcher, D., Gilchrist, G.S (2000). An evaluation of the impact of a self-care coping intervention on psychological and clinical outcomes in adolescents with newly diagnosed cancer. Eur J Oncol Nurs. 4(1):6-17; discussion 18-9. doi: 10.1054/ejon.1999.0051. PMID: 128496

[21] National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology, Cancer-Related Fatigue, Version 2.2022. NCCN, 2022

[22] Cella, D., Choi, S., Garcia, S., Cook, K.F., Rosenbloom, S., Lai, J-S, Tatum, D.S., Gershon, R. Setting Standards for Severity of Common Symptoms in Oncology Using the PROMIS Item Banks and Expert Judgement (2014). Qual Life Res.(10):2651-61. doi: 10.1007/s11136-014-0732-6.

MUC2023-190 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	Some data elements are in defined fields in electronic sources. Burden for providers was assessed across 9 sites via a questionnaire. Respondents reported the majority of the implementation burden was associated with administering the survey rather than collecting the clinical and demographic data elements; patient identification was also a challenge, which test sites mitigated by building EHR reports to facilitate patient identification. Alpha testing conducted July 1, 2019, to September 5, 2019, calculated eligible data elements, missing responses, invalid response options, and inappropriate answers. Results informed the 1) revision of data elements related to skipped, not applicable, and missing variables, 2) development of a REDCap form for Beta testing, 3) training of project managers and abstractors, and 4) quality control procedures. Beta testing was conducted from October 1, 2019, to March 31, 2021. Testing involved 20 sites; however, due to limited patient responses, data from 10 sites were analyzed. Beta test results informed specifications (e.g., numerator, denominator, denominator exclusions, and risk adjustment variables/model).	Manual abstraction could be required for some calculations.	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.

MUC2023-190 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden to data collection and reporting? (*Context of Use)*	Evidence provided demonstrated that performance measurement aimed at assessing provider accountability may help to drive improvements on the measure target. As specified, improvements on the measure benchmark will likely have significant impact on population outcomes.	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound? *(Context of Use)*	Signal-to-noise reliability analysis was conducted (n=22). Median result was 0.16. Hierarchical linear regression modeling was used to evaluate measure reliability. Clinician-level reliability testing estimated the adjusted ICC was 0.027. The reliability estimate with the average sample size for a clinician (8 patients per clinician) was 0.18. The Spearman-Brown prophecy formula was employed, and results indicated that an average sample size of 83 patient respondents was required to obtain a nominal reliability of 0.70. No clinicians in the sample had reliability at or above 0.70. Group-level reliability testing estimated the adjusted ICC was 0.094. The reliability estimate with the average sample size for a group (32 patients per group) was 0.77. The Spearman-Brown prophecy formula was used again, and results indicated an	The reliability for groups with less than 16 patients may be below 0.6.	Most or all entities have reliability above the threshold (0.60) *within* the quality program population.

MUC2023-190 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	average sample size of 23 patient respondents would be required for sufficient reliability (i.e., at or above 0.70). Group-specific reliability ranged from 0.38 to 0.88 (M=0.66; SD=0.21), and a median reliability of 0.68. In the sample, 50% of groups had reliability scores at or above 0.70. The performance score is the average patient-level score for each entity. To assess the reliability of the quality measure, a traditional "signal-to-noise" analysis is used that decomposes variability in the measure score into a) between-subject variability and b) within-subject variability. If there is a large amount of between-subject variability (i.e., "signal") compared to within-subject variability (i.e., "noise"), then there is more evidence that it is possible to discriminate performance among clinicians or groups. To evaluate quality measure reliability for clinician-level reporting, the developer used hierarchical linear regression models to relate outcome measures to providers and their covariates. The variance of the model can be decomposed using the adjusted intraclass correlation coefficient (ICC), which provides a summary of the reliability of the measure as tested, with higher values implying more variability between clinicians. The reliability from the measure test is then		

MUC2023-190 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	projected out based on observed variances and sample sizes from each clinician, using the Spearman-Brown prophecy formula. The measure report indicates a median signal-to-noise reliability of 0.16 for clinician level and 0.68 for group level. The reliability for groups with at least 16 patients may be above 0.6.		
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	SPSTF Grade A, Strong recommendation or similar. Empiric validity was also tested (n=7), yielding a result of -0.509. Further, to test empiric validity, the measure developer collected data from test sites during the testing time period for HCAHPS, Outpatient Oncology Press Ganey, and QOPI. Without viewing submitted data, TEP members rated expected correlation strength between the Patient-Reported Outcome Measures Oncology (PROMOnc) Survey measures and these available data. The measure developer then analyzed correlations for any measure for which the TEP hypothesized a moderate association and for which they had data for at least 7 test sites. The measure was moderately correlated with other patient-reported	Upon additional analysis, if the five votes reporting moderate agreement (e.g., "3" rating), merely three voters reported complete agreement that the fatigue measure could distinguish between good versus poor quality. Voters that rated the measure as less than 4 reported thoughts that fatigue could be susceptible to pandemic-related issues. Of note, four oncologists declined to vote regarding face validity voting. These oncologists noted concerns related to COVID-19's effect on sample size as well as performance scores.	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population.

MUC2023-190 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	outcome measures in the predicted direction, indicating the measure is associated with better reported patient experience. Face validity was assessed with eight voters. All voters reported moderate to total agreement on a scale from 1-5, indicating the measure can distinguish good from poor quality care among accountable entities.		
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	Patient-level risk adjustors were clearly defined (e.g., age, BMI at baseline, race and ethnicity). Specifically, 38 variables were identified and considered during testing, with 26 variables included in PROMOnc testing (e.g., demographics, social risk factors, comorbidities, baseline survey scores).	--	N/A
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	This measure has the opportunity for improvement at the clinician group level.	--	There is an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-190 Measure Reliability

The performance score is the average patient-level score for each entity.

The supplemental group level testing document indicates reliability at the average sample size for a group (32 patients per group) of 0.77. Simulated reliability tables were developed given the reliability method chosen and availability of data.

Interpretation:

The reported reliability at the group level is 0.77 at the average sample size of 32 patients per group. Using the Spearman-Brown prophecy formula, the reliability for groups with less than 16 patients is below 0.6.

3.5 MUC2023-205 Inpatient (IP) Percutaneous Coronary Intervention (PCI)

Description: The Inpatient (IP) Percutaneous Coronary Intervention (PCI) episode-based cost measure evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients who present with a cardiac event and emergently receive PCI as treatment.

Measure Type: Cost/Resource Use

Level of Analysis: Clinician Group

Data Source(s): Administrative Data (non-claims); Claims Data

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Inpatient (IP) Percutaneous Coronary Intervention (PCI) episode-based cost measure to the MIPS measure set. This measure evaluates a clinician's or clinician group's risk-adjusted and specialty-adjusted cost to Medicare for the urgent PCI treatment of patients who present with a cardiac event. The IP PCI measure is a revised version of a measure currently in use in MIPS that has been in the program since 2019, ST-Elevation Myocardial Infarction (STEMI) with Percutaneous Coronary Intervention (PCI). This revised IP PCI measure improves upon the original measure by greatly increasing the number of patients and clinicians covered by this measure. The substantive updates to this measure expand the patient cohort beyond beneficiaries treated with PCI for STEMI to include beneficiaries receiving a PCI with a Non-STEMI (NSTEMI) diagnosis and without a STEMI or NSTEMI diagnosis. Clinicians have opportunities to improve their care for patients who underwent PCI treatments, including reducing PCI complications, readmissions, and resulting acute kidney injury (AKI). CMS is considering this measure for use in MIPS because there are opportunities to improve patient outcomes and reduce the cost to Medicare for patients treated with IP PCI.

Table 3.5.1. MUC2023-205 Brief Summary of Measure Information

Merit Submission Information MUC2023-205	Description
Measure name	Inpatient (IP) Percutaneous Coronary Intervention (PCI)
MUC ID	MUC2023-205
Cascade priority	Affordability and Efficiency
Measure steward	Centers for Medicare & Medicaid Services
Measure Developer	Acumen, LLC

Merit Submission Information MUC2023-205	Description
Program submitted to	Merit-based Incentive Payment System-Cost
Committee assigned to	Clinician Committee
Related measures in the program	ST-Elevation Myocardial Infarction (STEMI) with Percutaneous Coronary Intervention (PCI) episode-based measure
Is this a new measure in this year's MUC list?	No
If not a new measure, then describe the history of this measure in prior MUC list inclusion	This measure was submitted as MUC17-262 to the Merit-based Incentive Payment System, 2017-2018 and was reviewed by the Clinician Workgroup, leading to a recommendation of conditional support.
Is the measure currently used in a CMS program?	Yes
If previously used, please describe the history of the measure in CMS program	Merit-based Incentive Payment System-Cost, 2019-2023
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	Section 101(f) of Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
Measure Specification Details	
Measure description	The Inpatient (IP) Percutaneous Coronary Intervention (PCI) episode-based cost measure evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients who present with a cardiac event and emergently receive PCI as treatment. This acute inpatient medical condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during each episode from the clinical event that opens, or "triggers," the episode through 30 days after the trigger.
Data source	Administrative Data (non-claims); Claims Data
Level of analysis	Clinician: Group only
Numerator	The measure numerator is the sum of the ratio of observed to expected payment-standardized cost to Medicare for all Inpatient (IP) Percutaneous Coronary Intervention (PCI) episodes attributed to a clinician. This sum is then multiplied by the national average observed episode cost to generate a dollar figure.
Denominator	The measure denominator is the total number of episodes from the Inpatient (IP) Percutaneous Coronary Intervention (PCI) episode group assigned to the clinician.

Merit Submission Information MUC2023-205	Description
Numerator exclusions	N/A
Denominator exclusions	The following standard exclusions are applied to ensure data completeness: • 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. • Patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window or was enrolled in Part C for any part of the lookback plus episode window. • No clinician group (identified by TIN) is attributed the episode. • Patient's date of birth is missing. • Patient's death date occurred before the episode ended. • The trigger IP stay has the same admission date as another IP stay. Exclusions specific to the Inpatient (IP) Percutaneous Coronary Intervention (PCI) measure are developed with input from the Inpatient (IP) Percutaneous Coronary Intervention (PCI) Clinician Expert Workgroup, and include recent hospitalization for STEMI, recent hospitalization for respiratory failure, or patients with new cardiac device implantation, history of intracranial hemorrhage or cerebral infarction, transplant, shock/cardiac arrest, or ventilator dependence.
Denominator exceptions	N/A
Risk adjustment	Yes
Development Status	Fully Developed
Target population	Medicare Fee-for-Service (FFS)
Measure type	Cost/Resource Use
Is the measure a composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A

Merit Submission Information MUC2023 205	Description
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.5.2. MUC2023-205 Measure Evaluation

MUC2023 205 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	Studies cited show relevance of cost containment related to inpatient percutaneous coronary intervention (PCI).[23,24] Evidence provided established that reducing complications related to bleeding, thrombosis, and ischemic events can improve patient outcomes and reduce downstream costs associated with subsequent treatment[25], aligning with clinical goals and priorities for coronary artery disease (CAD) patients.	--	The study population is the same as the target quality program population.
Conformance: Does the measure as specified align with the conceptual intent?	This measure's specifications align with the measure's intent. It has appropriate attribution at clinician group level and specification which includes observed cost to the scaled	--	Most persons and entities in the quality program population are included in the specification.

[23] Inohara T, Kohsaka S, Spertus JA, et al. Comparative Trends in Percutaneous Coronary Intervention in Japan and the United States, 2013 to 2017. *J Am Coll Cardiol*. 2020;76(11):1328-1340. doi:10.1016/j.jacc.2020.07.037

[24] Afana M, Brinjikji W, Cloft H, Salka S. Hospitalization costs for acute myocardial infarction patients treated with percutaneous coronary intervention in the United States are substantially higher than Medicare payments. *Clin Cardiol*. 2015;38(1):13-19. doi:10.1002/clc.22341.

[25] Lawton JS, Tamis-Holland JE, Bangalore S, et al. 2021 ACC/AHA/SCAI Guideline for Coronary Artery Revascularization: Executive Summary: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines [published correction appears in Circulation. 2022 Mar 15;145(11):e771]. *Circulation*. 2022;145(3):e4-e17. doi:10.1161/CIR.0000000000001039

MUC2023 205 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
(Concept of Interest)	expected cost for all PCI episodes attributed to a clinician, where each ratio is weighted by each episode's number of days assigned to a clinician.		Data element reliability and validity extrapolate to the quality program population.
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	This is a claims-based measure. It can be calculated exclusively from claims submitted electronically for billing and other purposes. All data elements are in defined fields in electronic sources.	No explicit articulation of people, processes, or technology required.	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden to data collection and reporting? (*Context of Use)*	Evidence provided demonstrated that performance measurement aimed at assessing provider accountability may help to drive improvements on the measure target. As specified, improvements on the measure benchmark will likely have significant impact on population outcomes such as episode-based cost but also potentially a reduction in harms and readmissions.	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound? *(Context of Use)*	Signal-to-noise reliability testing was conducted at the appropriate level. With a minimum of 30 episodes per entity, the median reliability is 0.69. About 85% of the entities may have reliability above 0.6.	With a threshold of at least 10 episodes per entity, the median reliability is 0.53. The reliability for about 55% of the entities is below 0.6. When the threshold is increased to a minimum of 20 episodes per entity, the median reliability is 0.63. The reliability for about 35% of the entities is below 0.6.	Some entities have reliability above the threshold (0.60) *within* the quality program population.

MUC2023 205 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
		When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.69. About 15% of the entities may have reliability below 0.6.	
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	Validity was evaluated empirically by estimating the effect of relevant treatment choices on the measure score. There was a strong association between measure focus and measured entity (clinician group). Clinician groups can assert reasonable influence and improve outcomes on this measure based on associations seen in empiric testing.	While there was not a clear plan for how an entity may improve performance on the measure focus, the treatment choice assessments during validity testing suggest that such a plan could easily be developed from available data.	There is an association between the entity and the measure focus *within* the quality program population.
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	This measure is risk adjusted and there is an explicit rationale for confounders included in the model. Risk adjustment uses CMS Hierarchical Condition Code Version 24 (CMS-HCC V24) 2021 Risk Adjustment Model. This includes comorbidities captured by 86 HCC codes that map with thousands of ICD-10-CM codes and other standard risk adjustors, including interaction variables accounting for a range of comorbidities, patient-level demographics such as age and health status (i.e., disability status, recent use of long-term care), HCC count, patient dual eligibility status, and number and	--	N/A

MUC2023 205 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	types of clinician specialties from which the patient has received care.		
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	This measure has the opportunity for improvement at the clinician group level. The measure's development is aligned with episode-based cost measures currently used in the program.	There is not an explicit articulation of the resources and context that might facilitate or be a barrier to the way an entity may improve.	There is an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-205 Measure Reliability[26]

The performance score is a risk- and specialty-adjusted average cost per episode.

Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across all entities. σ^2_{within} is the standard deviation of the score for a single entity.

For TIN-level entities with at least 10 episodes, the information provided indicates a median signal-to-noise reliability of 0.53 across 1,845 entities.

For TIN-level entities with at least 20 episodes, the information provided indicates a median signal-to-noise reliability of 0.63 across 1,202 entities.

For TIN-level entities with at least 30 episodes, the information provided indicates a median signal-to-noise reliability of 0.69 across 862 entities.

Decile tables:

[26] Developer provided additional decile tables to replace simulated reliability calculations.

The developer provided decile tables of performance score for the 20-episode threshold level and decile tables for reliability for the 10-, 20-, and 30-episode threshold levels.

For Table 3.5.3, entities are sorted by performance score, and the average score by decile is reported along with the number of entities and episodes included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.5.3. MUC2023-205 Performance Score Deciles –TINs with at least 20 attributed episodes

MUC2023-205	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	$20,468 (1,060)	16,968	19,238	19,589	19,900	20,159	20,357	20,598	20,895	21,255	21,796	23,526	26,309
Entities	1,202	1	120	120	120	120	120	120	120	120	120	120	1
# Episodes	69,639	--	--	--	--	--	--	--	--	--	--	--	--

For Tables 3.5.4, 3.3.5, and 3.5.6, entities are sorted by reliability, and the average reliability by decile is reported. Average and standard deviation are also included.

Table 3.5.4. MUC2023-205 Mean Reliability (by Reliability Decile) – TINs with at least 10 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.53	0.175	0.315	0.349	0.408	0.457	0.508	0.572	0.638	0.703	0.779	0.902

Table 3.5.4. MUC2023-205 Mean Reliability (by Reliability Decile) – TINs with at least 20 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.63	0.130	0.457	0.499	0.534	0.579	0.622	0.661	0.703	0.751	0.816	0.907

Table 3.5.5. MUC2023-205 Mean Reliability (by Reliability Decile) – TINs with at least 30 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.69	0.102	0.558	0.592	0.622	0.647	0.678	0.710	0.744	0.789	0.844	0.916

Interpretation:

With a threshold of at least 10 episodes per entity, the median reliability is 0.53. About 55% of the entities may have reliability below 0.6.

When the threshold is increased to a minimum of 20 episodes per entity, the median reliability is 0.63. About 35% of the entities may have reliability below 0.6.

When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.69. About 15% of the entities may have reliability below 0.6.

3.6 MUC2023-206 Kidney Transplant Management

Description: The Kidney Transplant Management 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 related to kidney transplant, beginning 90-days post-transplant.

Measure Type: Cost/Resources Use

Level of Analysis: Clinician Group

Data Source(s): Administrative Data (non-claims); Claims Data; Registries

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Kidney Transplant Management episode-based cost measure to the MIPS measure set. The measure assesses a clinician's or clinician group's Medicare costs for patients' post-kidney transplant. Increasing access to kidney transplants and maintaining excellent, high-value post-transplant care are priority areas for CMS. The Kidney Transplant Management measure, in conjunction with the Chronic Kidney Disease and Kidney Transplant Management Disease measures that CMS is also considering, will fill a measurement gap area in the care continuum for a patient with kidney disease as there are currently no episode-based cost measures for this clinical area. CMS is considering this measure for use in MIPS because there are opportunities to improve patient outcomes and to reduce the cost to Medicare for managing patients' care following a kidney transplant, such as reducing downstream costs of readmissions and emergency department visits or encouraging medication adherence.

Table 3.6.1. MUC2023-206 Brief Summary of Measure Information

MERIT Submission Information MUC2023-206	Description
Measure name	Kidney Transplant Management
MUC ID	MUC2023-206
Cascade priority	Affordability and Efficiency
Measure steward	Centers for Medicare & Medicaid Services
Measure developer	Acumen, LLC
Program submitted to	Merit-based Incentive Payment System-Cost

MERIT Submission Information MUC2023-206	Description
Committee assigned to	Clinician Committee
Related measures in the program	N/A
Is this a new measure in this year's MUC list?	Yes
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program	No
If previously used, please describe the history of the measure in CMS program	N/A
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	Section 101(f) of Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
Measure specification details	
Measure description	The Kidney Transplant Management 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 related to kidney transplant, beginning 90-days post-transplant. This chronic condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during a Kidney Transplant Management episode.
Data source	Administrative Data (non-claims); Claims Data; Registries
Level of analysis	Clinician: Group only
Numerator	The measure numerator is the weighted average ratio of the winsorized scaled standardized observed cost to the scaled expected cost for all Kidney Transplant Management episodes attributed to a clinician, where each ratio is weighted by each episode's number of days assigned to a clinician. This sum is then multiplied by the national average winsorized scaled observed episode cost to generate a dollar figure.
Denominator	The measure denominator is the total number of days from Kidney Transplant Management episodes assigned to the clinician across all patients.
Numerator exclusions	N/A

MERIT Submission Information MUC2023-206	Description
Denominator exclusions	The following standard exclusions are applied to ensure data completeness: • Patient has a primary payer other than Medicare for any time overlapping the episode window or 365-day lookback period prior to the episode window. • Patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window or was enrolled in Part C for any part of the lookback plus episode window. • Patient was not found in the Medicare Enrollment Database (EDB). • Patient's death date occurred before the episode end date. • Patient has an episode window shorter than 1 year. • Patients with extremely low treatment costs. • Patient's residence is outside the United States or its territories during the episode window, as indicated in the EDB. Exclusions specific to the Kidney Transplant Management measure are developed with input from the Kidney Transplant Management Clinician Expert Workgroup and include atypical hemolytic-uremic syndrome (HUS) and prior organ transplant for heart, intestine, liver, lung, or pancreas.
Denominator exceptions	N/A
Risk adjustment	Yes
Development Status	Fully Developed
Target population	Medicare Fee-for-Service (FFS)
Measure type	Cost/Resource Use
Is the measure a composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.6.2. MUC2023-206 Measure Evaluation

MUC2023-206 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	Studies cited show relevance of cost containment related to kidney transplant in the areas of 1) readmissions and emergency department (ED) visits, 2) immunosuppression regimens and medication adherence to prevent kidney rejection, 3) management of comorbidities including cardiovascular disease (CVD) management, and 4) infection prevention and control. Implementing a value-based approach to kidney transplant more broadly may help incentivize clinicians better to reduce post-transplant readmissions, medications, and delayed graft function or graft failure, which contribute significantly to overall health care costs for patients with a kidney transplant.[27,28,29]	No evidence submitted directly showing mechanisms by which clinician groups will improve kidney transplant outcomes through implementation of cost containment measures.	The study population is the same as the target quality program population
Conformance:	This measure's specifications align with the measure's intent. It has appropriate attribution at the clinician	--	Most persons and entities in the quality program population are included in the specification.

[27] Famure O, Kim ED, Au M, et al. What Are the Burden, Causes, and Costs of Early Hospital Readmissions After Kidney Transplantation? *Prog Transplant*. 2021;31(2):160-167. doi:10.1177/15269248211003563.
[28] Lenihan CR, Liu S, Airy M, Walther C, Montez-Rath ME, Winkelmayer WC. The Association of Pre-Kidney Transplant Dialysis Modality with de novo Posttransplant Heart Failure. *Cardiorenal Med*. 2021;11(5-6):209-217. doi:10.1159/000518535.
[29] Sussell J, Silverstein AR, Goutam P, et al. The economic burden of kidney graft failure in the United States. *Am J Transplant*. 2020;20(5):1323-1333. doi:10.1111/ajt.15750.

MUC2023-206 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	group level and includes in specification observed cost to the scaled expected cost for all kidney transplant management episodes attributed to a clinician, where each ratio is weighted by each episode's number of days assigned to a clinician.		
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	This is a claims-based measure. It can be calculated exclusively from claims submitted electronically for billing and other purposes. All data elements are in defined fields in electronic sources.	No explicit articulation of people, processes, or technology required.	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden to data collection and reporting? (*Context of Use)*	Evidence provided demonstrated that performance measurement aimed at assessing provider accountability may help to drive improvements on the measure target. As specified, improvements on the measure benchmark will likely have significant impact on population outcomes such as episode-based cost but also potentially quality of life.	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound? *(Context of Use)*	Appropriate signal-to-noise reliability testing was conducted in a relevant study population.	With a threshold of at least 10 episodes per entity, the median reliability is 0.291. The reliability for about 75% of the entities is below 0.6. When the threshold is increased to a minimum of 20 episodes per entity, the median reliability is	Some entities have reliability above the threshold (0.60) *within* a population that can be extrapolated to the program population.

MUC2023-206 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
		0.393. The reliability for about 65% of the entities is below 0.6. When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.468. About 60% of the entities may have reliability below 0.6.	
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	Validity was evaluated empirically by estimating the effect of relevant treatment choices on the measure score. There was a strong association between measure focus and measured entity (clinician group). Clinician groups can assert reasonable influence and improve outcomes on this measure based on associations seen in empiric testing.	While there was not a clear plan for the explicit articulation of the way an entity may improve performance on the measure focus, the treatment choice assessments during validity testing suggest that such a plan could easily be developed from data available.	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population.
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	This measure is risk adjusted, and there is an explicit rationale for confounders included in the model. Risk adjustment uses CMS Hierarchical Condition Code Version 24 (CMS-HCC V24) 2021 Risk Adjustment Model. This includes comorbidities captured by 86 HCC codes that map with thousands of ICD-10-CM codes and other standard risk adjustors, including interaction variables accounting for a range of comorbidities, patient-level demographics (i.e., age) and health status (i.e., disability status, end-stage renal disease [ESRD] status, recent use of long-term care), HCC count,	--	N/A

MUC2023-206 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	patient dual eligibility status, and number and types of clinician specialties from which the patient has received care.		
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	This measure has the opportunity for improvement at the clinician group level. The measure's development is aligned with episode-based cost measures currently used in the program. The Kidney Transplant Management measure was also developed in consideration of alignment opportunities with CMS' KCF and CKCC payment Options of the KCC Advanced Payment Model.	There is not an explicit articulation of the resources and context that might facilitate or be a barrier to the way an entity may improve.	There is an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-206 Measure Reliability[30]

The performance score is a risk- and specialty-adjusted average cost per episode. Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across all entities. σ^2_{within} is the standard deviation of the score for a single entity.

For TIN-level entities with at least 10 episodes, the information provided indicates a median signal-to-noise reliability of 0.291.

For TIN-level entities with at least 20 episodes, the information provided indicates a median signal-to-noise reliability of 0.393.

For TIN-level entities with at least 30 episodes, the information provided indicates a median signal-to-noise reliability of 0.468.

Decile tables:

[30] Developer provided additional decile tables to replace simulated reliability calculations.

The developer provided decile tables of performance score for the 20-episode threshold level and decile tables for reliability for the 10-, 20-, and 30-episode threshold levels.

For Tables 3.6.3, entities are sorted by performance score, and the average score by decile is reported along with the approximate number of entities included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.6.3. MUC2023-206 Performance Score Deciles – TINs with at least 20 attributed episodes

MUC2023-206	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	20,801 (4,255)	9,706	15,722	17,235	18,505	19,587	20,556	21,547	22,720	23,965	26,188	33,010	39,433
Entities	696	1	70	70	70	70	70	70	70	70	70	70	1

For Tables 3.6.4, 3.6.5, and 3.6.6, entities are sorted by reliability, and the average reliability by decile is reported. Average and standard deviation are also included.

Table 3.6.4. Mean Reliability (by Reliability Decile) – TINs with at least 10 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.331	0.208	0.092	0.142	0.187	0.233	0.291	0.349	0.416	0.507	0.650	0.842

Table 3.6.5. MUC2023-206 Mean Reliability (by Reliability Decile) – TINs with at least 20 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.423	0.201	0.177	0.232	0.290	0.343	0.393	0.454	0.528	0.621	0.735	0.859

Table 3.6.6. MUC2023-206 Mean Reliability (by Reliability Decile) – TINs with at least 30 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.486	0.191	0.249	0.308	0.360	0.415	0.467	0.526	0.608	0.673	0.761	0.866

Interpretation:

With a threshold of at least 10 episodes per entity, the median reliability is 0.291. About 75% of the entities may have reliability below 0.6.

When the threshold is increased to a minimum of 20 episodes per entity, the median reliability is 0.3934. About 65% of the entities may have reliability below 0.6.

When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.468. About 60% of the entities may have reliability below 0.6.

3.7 MUC2023-207 Prostate Cancer

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.

Measure Type: Cost/Resources Use

Level of Analysis: Clinician Group

Data Source(s): Administrative Data (non-claims); Claims Data; Registries

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Prostate Cancer episode-based cost measure to the MIPS measure set. This measure evaluates a clinician's or clinician group's risk-adjusted and specialty-adjusted cost to Medicare for the management and treatment of prostate cancer. The Prostate Cancer measure will fill a measurement gap area in care associated with cancer in the MIPS cost performance category, as there is currently no episode-based cost measure for this clinical area. CMS is considering this measure for use in MIPS because it has the potential to capture a large number of patients and Medicare spending and because there are opportunities for clinicians to improve their performance, such as reducing overtreatment among certain patient populations and overuse of monitoring tests for patients under active surveillance.

Table 3.7.1. MUC2023-207 Brief Summary of Measure Information

MERIT Submission Information MUC2023-207	Description
Measure name	Prostate Cancer
MUC ID	MUC2023-207
Cascade priority	Affordability and Efficiency
Measure steward	Centers for Medicare & Medicaid Services
Measure developer	Acumen, LLC
Program submitted to	Merit-based Incentive Payment System-Cost
Committee assigned to	Clinician Committee

MERIT Submission Information MUC2023-207	Description
Related measures in the program	N/A
Is this a new measure in this year's MUC List?	Yes
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program?	N/A
If previously used, please describe the history of the measure in CMS program	New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	Section 101(f) of Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
Measure Specification Details	
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.
Data source	Administrative Data (non-claims); Claims Data
Level of analysis	Clinician: Group only
Numerator	The measure numerator is the weighted average ratio of the winsorized scaled standardized observed cost to the scaled expected cost for all Prostate Cancer episodes attributed to a clinician, where each ratio is weighted by each episode's number of days assigned to a clinician. This sum is then multiplied by the national average winsorized scaled observed episode cost to generate a dollar figure.
Denominator	The measure denominator is the total number of days from Prostate Cancer episodes assigned to the clinician across all patients.
Numerator exclusions	N/A
Denominator exclusions	The following standard exclusions are applied to ensure data completeness: • Patient has a primary payer other than Medicare for any time overlapping the episode window or 365-day lookback period prior to the episode window.

MERIT Submission Information MUC2023-207	Description
	• Patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window or was enrolled in Part C for any part of the lookback plus episode window. • Patient was not found in the Medicare Enrollment Database (EDB). • Patient's death date occurred before the episode end date. • Patient has an episode window shorter than 1 year. • Patients with extremely low treatment costs. • Patient's residence is outside the United States or its territories during the episode window, as indicated in the EDB. Exclusions specific to the Prostate Cancer measure are developed with input from the Prostate Cancer Clinician Expert Workgroup, and include patients with hospice use in the past 1 year.
Denominator exceptions	N/A
Risk adjustment	Yes
Development status	Fully Developed
If not fully developed, development stage	N/A
Target population	Medicare Fee-for-Service (FFS)
Measure type	Cost/Resource Use
Is the measure a composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.7.2. MUC2023-207 Measure Evaluation

MUC2023-207 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	Prostate cancer is the second most common cancer among men in the United States.[31] In 2020, 9.3% of male Medicare beneficiaries had a prostate cancer diagnosis, slightly increasing from 8.9% in 2017.[32] 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.[33] In 2020, annual prostate cancer spending was between $18-19 billion per year, increasing faster than other cancer types.[34,35]	--	The study population is the same as the target quality program population
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	No empirical evidence of conformance was submitted. The cohort for this measure consists of Medicare beneficiaries enrolled in	--	Most persons and entities in the quality program population are included in the specification.

[31] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7-30. doi:10.3322/caac.21590.
[32] Chronic Conditions Data Warehouse. "Medicare chronic conditions charts." ccwdata.org (2021). Medicare Chronic Condition Charts - Chronic Conditions Data Warehouse (ccwdata.org)
[33] 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.
[34] 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.
[35] 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

MUC2023-207 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	Medicare fee-for-service who receive medical care to manage and treat prostate cancer. There are measure-specific exclusions including recent hospice use as well as standard exclusions to ensure data completeness. The Prostate Cancer measure focuses on the care provided by clinicians practicing in non-inpatient hospital settings for patients with prostate cancer. The most frequent settings in which a prostate cancer episode is triggered include office and outpatient hospital settings.		
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	No explicit articulation of people, processes, or technology required for measurement was submitted. This is a claims-based measure. It can be calculated from claims and other data submitted electronically for billing and other purposes. All data elements are in defined fields in electronic sources.	--	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden of data collection and reporting? (*Context of Use)*	Existing literature and the measure developer's preliminary testing indicate a high cost to Medicare for treating and managing prostate cancer, opportunities for improvement through best practices, and a substantial empirical performance gap. Testing indicates that the Prostate Cancer	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.

MUC2023-207 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	measure would significantly impact on Medicare, measured by cost and beneficiary count. Currently, the mean scores are $11,480.14 and $10,947.21 by TIN and TIN-NPI respectively, with SDs of $3,846.80 and $4,430.54, respectively.		
Reliability: Is measure performance scientifically sound? *(Context of Use)*	Signal-to-noise reliability testing was conducted at the appropriate level. The measure developer reports that at the testing volume of 20 episodes, the mean reliability for the Prostate Cancer measure is 0.68 and 0.61 at the TIN and TIN-NPI level, respectively. With a minimum of 30 episodes per entity, the median reliability is 0.758. About 85% of the entities may have reliability above 0.6.	With a threshold of at least 10 episodes per entity, the median reliability is 0.654. About 35% of the entities may have reliability below 0.6. When the threshold is increased to a minimum of 20 episodes per entity, the median reliability is 0.715. About 25% of the entities may have reliability below 0.6. When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.758. About 15% of the entities may have reliability below 0.6.	Some entities have reliability above the threshold (0.60) within a population that can be extrapolated to the program population.
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	Over-screening and aggressive treatments for older adults with prostate cancer are less beneficial. Previous studies found no significant survival benefit within 10 years for	--	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population.

MUC2023-207 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	screening in older men.[36, 37, 38, 39] There is increasing evidence that certain combinations of testing methodologies can provide clinicians with more accurate staging and diagnostic information, conserve valuable testing resources, and improve QOL and patient outcomes during treatment.[40]		
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	The materials submitted do not provide a detailed rationale for confounders included in the model. However, the measure developer notes they received generalized feedback on risk adjustment in episode-based cost measure calculation during a previous TEP meeting. The Prostate Cancer Clinician Expert Workgroup's feedback on risk adjustors was incorporated into the model. The model is stratified by Part D enrollment status. Risk adjustment is performed separately	--	N/A

[36] 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
[37] 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.
[38] 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.
[39] 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.
[40] Russo F, Mazzetti S, Regge D, et al. Diagnostic Accuracy of Single-plane Biparametric and Multiparametric Magnetic Resonance Imaging in Prostate Cancer: A Randomized Noninferiority Trial in Biopsy-naïve Men. Eur Urol Oncol. 2021;4(6):855-862. doi:10.1016/j.euo.2021.03.007.

MUC2023-207 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	within each stratification. The draft measure also underwent a national field-testing period and public comment periods, after which the Clinical Expert Workgroup further refined measure specifications based on feedback collected during testing. The risk adjustment model for this measure uses a linear regression model, which utilizes variables from the CMS Hierarchical Condition Code Version 24 (CMS-HCC V24) 2021 Risk Adjustment Model. This includes comorbidities captured by 86 HCC codes that map with thousands of ICD-10-CM codes, interaction variables accounting for a range of comorbidities, and other standard risk adjustors, including patient-level demographics (i.e., age) and health status (i.e., disability status, end-stage renal disease [ESRD] status, recent use of long-term care).		
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	--	The submitted materials do not include an explicit articulation of the resources and context that might facilitate or be a barrier to the way an entity may improve.	Unable to determine if there is an articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-207 Measure Reliability[41]

The performance score is a risk- and specialty-adjusted average cost per episode.

Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across all entities. σ^2_{within} is the standard deviation of the score for a single entity.

For TIN-level entities with at least 10 episodes, the information provided indicates a median signal-to-noise reliability of 0.654 across 4,734 entities.

For TIN-level entities with at least 20 episodes, the information provided indicates a median signal-to-noise reliability of 0.715 across 3,067 entities.

For TIN-level entities with at least 30 episodes, the information provided indicates a median signal-to-noise reliability of 0.758 across 2,438 entities.

Decile tables:

The developer provided decile tables of performance score for the 20-episode threshold level and decile tables for reliability for the 10-, 20-, and 30-episode threshold levels.

For Tables 3.7.3, entities are sorted by performance score, and the average score by decile is reported along with the number of entities and episodes included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.7.3. MUC2023-207 Performance Score Deciles – TINs with at least 20 attributed episodes

MUC2023-207	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	$11,480 ($3,847)	$2,244	$6,895	$8,358	$9,364	$10,398	$11,181	$12,037	$12,989	$14,263	$16,174	$23,155	$31,191
Entities	3,067	1	307	307	307	307	307	307	307	307	307	473307	1
# Episodes	555,178 10,21	--	--	--	--	--	--	--	--	--	--	--	--

[41] Developer provided additional decile tables to replace simulated reliability calculations.

For Tables 3.7.4, 3.7.5, and 3.7.8, entities are sorted by reliability, and the average reliability by decile is reported. Average and standard deviation are also included.

Table 3.7.4. MUC2023-207 Mean Reliability (by Reliability Decile) – TINs with at least 10 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.61	0.25	0.224	0.363	0.476	0.579	0.654	0.728	0.794	0.850	0.914	0.977

Table 3.7.5. MUC2023-207 Mean Reliability (by Reliability Decile) – TINs with at least 20 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.68	0.21	0.365	0.484	0.585	0.653	0.715	0.775	0.825	0.868	0.919	0.977

Table 3.7.6. MUC2023-207 Mean Reliability (by Reliability Decile) – TINs with at least 30 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.72	0.18	0.450	0.570	0.64	0.701	0.758	0.803	0.843	0.881	0.928	0.979

Interpretation:

With a threshold of at least 10 episodes per entity, the median reliability is 0.654. About 35% of the entities may have reliability below 0.6.

When the threshold is increased to a minimum of 20 episodes per entity, the median reliability is 0.715. About 25% of the entities may have reliability below 0.6.

When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.758. About 15% of the entities may have reliability below 0.6.

3.8 MUC2023-208 Respiratory Infection Hospitalization

Description: The Respiratory Infection Hospitalization episode-based cost measure evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients who receive inpatient treatment for a respiratory infection.

Measure Type: Cost/Resource Use

Level of Analysis: Clinician Group

Data Source(s): Administrative Data (non-claims); Claims Data

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Respiratory Infection Hospitalization episode-based cost measure to the MIPS measure set. This measure evaluates a clinician's or clinician group's risk-adjusted and specialty-adjusted cost to Medicare for the inpatient treatment of respiratory infection. This measure is a revised version of a measure that has been in the MIPS program since 2019, Simple Pneumonia with Hospitalization. In the CY 2024 Physician Fee Schedule (PFS) Final Rule, CMS finalized the removal of this measure from MIPS due to coding changes that unevenly impacted clinicians and prevented the measure from assessing pneumonia hospitalizations as intended. This revised Respiratory Infection Hospitalization measure addresses these concerns and improves upon the original measure by increasing the number of clinicians and patients covered by this measure. The substantive updates to this measure include expanding the patient cohort to include beneficiaries hospitalized for pneumonia and related respiratory infections not otherwise captured under the measure due to recent changes in coding guidance. Clinicians have opportunities to improve their care for patients with respiratory infection hospitalizations, including reducing hospital readmissions and overuse of antibiotics. CMS is considering this measure for use in MIPS because there are opportunities to improve patient outcomes and reduce the cost to Medicare for patients with respiratory infection hospitalizations.

Table 3.8.1. MUC2023-208 Brief Summary of Measure Information

MERIT Submission Information MUC2023-208	Description
Measure name	Respiratory Infection Hospitalization
MUC ID	MUC2023-208*
Cascade priority	Affordability and Efficiency

MERIT Submission Information MUC2023-208	Description
Measure steward	Centers for Medicare & Medicaid Services
Measure developer	Acumen, LLC
Program submitted to	Merit-based Incentive Payment System-Quality
Committee assigned to	Clinician Committee
Related measures in the program	The Simple Pneumonia with Hospitalization episode-based measure is currently used in the Merit-Based Incentive Payment System with the measure ID of COST_SPH_1 and is a previous iteration of the Respiratory Infection Hospitalization measure.
Is this a new measure in this year's MUC List?	No
If not a new measure, then describe the history of this measure in prior MUC list inclusion	This measure was submitted during MUC year 2017 as MUC17-365 to the 2017-2018 Merit-based Incentive Payment System and underwent review by the MAP Clinician Workgroup, leading to a supportive recommendation.
Is the measure currently used in a CMS program?	Yes
If previously used, please describe the history of the measure in CMS program	Measure currently used in the Merit-based Incentive Payment System-Cost (2019-prersent), but the measure is undergoing substantial change
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	Section 101(f) of Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
History of endorsement	N/A
Path to endorsement	Unknown
Measure Specification Details	
Measure Description	The Respiratory Infection Hospitalization episode-based cost measure evaluates a clinician's or clinician group's risk-adjusted cost to Medicare for patients who receive inpatient treatment for a respiratory infection. This acute inpatient medical condition measure includes the costs of services that are clinically related to the attributed clinician's role in managing care during each episode from the clinical event that opens, or "triggers," the episode through 30 days after the trigger.
Data source	Administrative Data (non-claims); Claims Data
Level of analysis	Clinician: Group only

MERIT Submission Information MUC2023-208	Description
Numerator	The measure numerator is the sum of the ratio of observed to expected payment-standardized cost for all Respiratory Infection Hospitalization episodes attributed to a clinician. This sum is then multiplied by the national average winsorized observed episode cost to generate a dollar figure.
Denominator	The measure denominator is the total number of episodes from the Respiratory Infection Hospitalization episode group assigned to the clinician.
Numerator exclusions	N/A
Denominator exclusions	The following standard exclusions are applied to ensure data completeness: • 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. • Patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window or was enrolled in Part C for any part of the lookback plus episode window. • No clinician group (identified by TIN) is attributed the episode. • Patient's date of birth is missing. • Patient's death date occurred before the episode ended. • The trigger IP stay has the same admission date as another IP stay. Exclusions specific to the Respiratory Infection Hospitalization measure are developed with input from the Respiratory Infection Hospitalization Clinician Expert Workgroup, and include pleurisy diagnosis, pleural conditions, pleural plaque, chest trauma, chest wall myopathy, epidemic myalgia, fibrothorax, influenza due to avian flu, adverse effects of glucocorticoids, hospitalizations for certain non-pneumonia diagnoses, and patients discharged against medical advice.
Denominator exceptions	N/A
Risk adjustment	Yes
Development Status	Fully Developed
If not fully developed, development stage	N/A
Target population	Medicare Fee-for-Service (FFS)
Measure type	Cost/Resource Use
Is the measure a composite or component of a composite?	No
Digital Measure Information	

MERIT Submission Information MUC2023-208	Description
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.8.2. MUC2023-208 Measure Evaluation

MUC2023-208 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	In 2015, nearly 6.8 million episodes of clinical pneumonia resulted in hospital admissions in older adults.[42] There are nearly 140,000 hospital readmissions and more than $10 billion in hospital expenditures related to pneumonia each year.[43] Transitional care interventions have been shown to reduce hospital readmissions for pneumonia.[44]	--	The study population is the same as the target quality program population

[42] Shi, Ting, Angeline Denouel, Anna K. Tietjen, Jen Wei Lee, Ann R. Falsey, Clarisse Demont, Bryan O. Nyawanda, et al. "Global and Regional Burden of Hospital Admissions for Pneumonia in Older Adults: A Systematic Review and Meta-Analysis." The Journal of Infectious Diseases 222, Oct 07, 2020. Pages S570–S576, https://doi.org/10.1093/infdis/jiz053.
[43] Alba, Israel De and Alpesh Amin. "Pneumonia Readmissions: Risk Factors and Implications." The Ochsner Journal 14, no. 4 (12, 2014): 649-654. https://www.proquest.com/scholarly-journals/pneumonia-readmissions-risk-factors-implications/docview/2157950821/se-
[44] Transitional Care Reduces Pneumonia Readmissions. Stanford Health Care. November 26, 2016. Accessed May 18, 2023. https://stanfordhealthcare.org/content/dam/SHC/clinics/aging-adult-services/docs/10.31.16%20MedStaff%20Update.pdf.

MUC2023-208 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	The average length of antibiotic therapy for community-acquired pneumonia often exceeds national recommendations.[45] One study found that patients with pneumonia who received antibiotic stewardship program interventions were less likely to be readmitted and lower rates of antibiotic expenditure.[46]		
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	The measure developer's testing indicates the Respiratory Infection Hospitalization cost measure would capture over 300,000 beneficiaries.	The materials submitted did not include empirical evidence of conformance.	Most persons and entities in the quality program population are included in the specification.
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	This is a claims-based measure. It can be calculated from claims and other data submitted electronically for billing and other purposes. All data elements are in defined fields in electronic sources.	No explicit articulation of people, processes, or technology required for measurement was submitted.	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed	Data submitted by the measure developer show variation in the measure score at both TIN and TIN-NPI levels. At the TIN level, the mean score is $15,066, with a standard deviation of $1,253. The 10th	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.

[45] Flanagan, Jane, Kelly D. Stamp, Matt Gregas, and Judy Shindul-Rothschild. "Predictors of 30-Day readmission for pneumonia." The Journal of Nursing Administration 46, no. 2 (2016): 69-74.
[46] Mauro, James, Saman Kannangara, Joanne Peterson, David Livert, and Roman A. Tuma. "Rigorous Antibiotic Stewardship in the Hospitalized Elderly Population: Saving Lives and Decreasing Cost of Inpatient Care." JAC-Antimicrobial Resistance 3, no. 3 (09, 2021): 1. https://doi.org/10.1093/jacamr/dlab118.

MUC2023-208 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
the burden to data collection and reporting? *(Context of Use)*	percentile score is $13,527, the 50th percentile score is $14,959, and the 90th percentile score is $16,767. At the TIN-NPI level, the mean score is $17,207, with a standard deviation of $1,893. The 10th percentile score is $14,917, the 50th percentile score is $17,061, and the 90th percentile score is $19,739.		
Reliability: Is measure performance scientifically sound? *(Context of Use)*	Reliability testing was conducted by the measure developer for individual clinicians (TIN-NPIs) and constructed using episodes ending between January 1, 2022, and December 31, 2022. With a minimum of 30 episodes per entity, the median reliability is 0.79. Nearly all entities may have a reliability above 0.6.	According to the decile table, over 50% of the entities have reliability below 0.6. When the threshold is increased to a minimum of 20 episodes per entity, the median reliability is 0.74. About 15-20% of the entities may have reliability below 0.6.	Some entities have reliability above the threshold (0.60) within a population that can be extrapolated to the program population.
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	Empirical testing results submitted by the measure demonstrate that the cost measure reflects cost directly related to treatment choices and adverse events. Costs associated with outpatient evaluation and management services and imaging are associated with worse measure scores. These activities appear to co-occur with adverse events, suggesting that reducing adverse events could reduce spending on these services. While physician services are associated with worse measure scores, they are also	Face validity testing was not conducted for the Respiratory Infection Hospitalization measure. The materials submitted did not include an explicit articulation of the way an entity may improve performance on the measure focus.	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population.

MUC2023-208 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	associated with lower costs of adverse events at the TIN-NPI level. This suggests additional costs for physician services, inpatient services during trigger hospitalization, and durable medical equipment costs could be offset by reducing costs of adverse events.		
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	The materials submitted do not provide a detailed rationale for confounders included in the model. However, the measure developer notes they received generalized feedback on risk adjustment in episode-based cost measure calculation during a previous TEP meeting. The Clinician Expert Workgroup's feedback on risk adjustors and exclusions was incorporated into the model. The draft measure also underwent a national field-testing period and public comment periods, after which the Clinical Expert Workgroup further refined measure specifications based on feedback collected during testing. The risk adjustment model for this measure uses a linear regression model, which utilizes variables from the CMS Hierarchical Condition Code Version 24 (CMS-HCC V24) 2021 Risk Adjustment Model. This includes comorbidities captured by 86 HCC codes that map with thousands of ICD-10-CM codes, interaction variables	--	N/A

MUC2023-208 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	accounting for a range of comorbidities, and other standard risk adjustors, including patient-level demographics (i.e., age) and health status (i.e., disability status, end-stage renal disease [ESRD] status, recent use of long-term care). Measure-specific risk adjustors were identified by the Respiratory Infection Hospitalization Clinical Expert Workgroup and include asthma, acid-base disorders, COVID-19, pleural effusion/thoracentesis, dementia, limited mobility, recent use of long-term assisted care within 30 days, recent all-cause admission in prior 120 days, and prior oxygen use/respiratory failure.		
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	Adherence to standard guidelines to care for Medicare beneficiaries hospitalized with pneumonia has been shown to decrease readmission rates.[47] In a study of hospitalized patients with lower respiratory infections, use of Procalcitonin (PCT), a biomarker that has shown promising results in guiding	The submitted materials do not include an explicit articulation of the resources and context that might facilitate or be a barrier to the way an entity may improve.	There is an explicit articulation of the resources and context that might facilitate improvement *within* the quality program population.

[47] Dean NC, Bateman KA, Donnelly SM, Silver MP, Snow GL, Hale D. Improved clinical outcomes with utilization of a community-acquired pneumonia guideline. Chest. 2006;130(3):794-799. doi:10.1378/chest.130.3.794.

MUC2023-208 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	antibiotic therapy.[48] LRTIs reduced total costs by $2,867, a difference driven by a reduction in patient length of stay and antibiotic resistance.[49]		

MUC203-208 Measure Reliability[50]

The performance score is a risk- and specialty-adjusted average cost per episode. Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across all entities. σ^2_{within} is the standard deviation of the score for a single entity.

For TIN-level entities with at least 10 episodes, the measure report indicates a median signal-to-noise reliability of 0.62 across 4,945 entities.

For TIN-level entities with at least 20 episodes, the measure report indicates a median signal-to-noise reliability of 0.74 across 3,169 entities.

For TIN-level entities with at least 30 episodes, the measure report indicates a median signal-to-noise reliability of 0.79 across 2,451 entities.

Decile tables:

The developer provided decile tables of performance score for the 20-episode threshold level and decile tables for reliability for the 10-, 20-, and 30-episode threshold levels.

[48] Falcone, Marco, Michael Bauer, Ricard Ferrer, Gaëtan Gavazzi, Juan Gonzalez Del Castillo, Alberto Pilotto, and Philipp Schuetz. "Biomarkers for Risk Stratification and Antibiotic Stewardship in Elderly Patients." Aging Clinical and Experimental Research (Mar 30, 2023). https://doi.org/10.1007/s40520-023-02388-w.

[49] Mewes, Janne C., Michael S. Pulia, Michael K. Mansour, Michael R. Broyles, H. B. Nguyen, and Lotte M. Steuten. "The Cost Impact of PCT-Guided Antibiotic Stewardship Versus Usual Care for Hospitalised Patients with Suspected Sepsis Or Lower Respiratory Tract Infections in the US: A Health Economic Model Analysis." PloS One 14, no. 4 (2019): 1. https://doi.org/10.1371/journal.pone.0214222

[50] Developer provided decile values to replace simulated reliability tables.

For Tables 3.8.3 and 3.8.4, entities are sorted by performance score, and the average score by decile is reported along with the number of entities and episodes included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.8.3. MUC2023-208 Performance Score Deciles – TINs with at least 20 attributed episodes

MUC2023-208	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	$15,066 (1,352)	10,424	13,527	14,025	14,360	14,652	14,959	15,268	15,601	16,032	16,767	18,949	21,670
Entities	3,169	1	317	317	317	317	317	317	317	317	317	317	1
# Episodes	353,164	--	--	--	--	--	--	--	--	--	--	--	--

For Tables 3.8.4, 3.8.5, and 3.8.6, entities are sorted by reliability and the average reliability by decile is reported. Average and standard deviation are also included.

Table 3.8.4. MUC2023-208 Mean Reliability (by Reliability Decile) – TINs with at least 10 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.62	0.199	0.356	0.413	0.429	0.460	0.525	0.593	0.673	0.834	0.903	0.976

Table 3.8.5. MUC2023-208 Mean Reliability (by Reliability Decile) – TINs with at least 20 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.74	0.143	0.536	0.584	0.637	0.683	0.734	0.788	0.840	0.886	0.927	0.979

Table 3.8.6. MUC2023-208 Mean Reliability (by Reliability Decile) – TINs with at least 30 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.79	0.111	0.637	0.673	0.711	0.757	0.796	0.835	0.873	0.904	0.938	0.981

Interpretation:

With a threshold of at least 10 episodes per entity, the reported median reliability is 0.62, however according to the decile table, the median is lower (about 0.53-0.57). According to the decile table, over 50% of the entities have reliability below 0.6. When the threshold is increased to a minimum of 20 episodes per entity, the median reliability is 0.74. About 15-20% of the entities may have reliability below 0.6. When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.79. Very few, or perhaps none of the entities have reliability below 0.6. This measure has a good capability of distinguishing the quality of performance between entities with at least 30 episodes. Methods may need to be considered to mitigate entities with fewer than 30 episodes.

3.9 MUC2023-209 Rheumatoid Arthritis

Description: The Rheumatoid Arthritis 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 rheumatoid arthritis.

Measure Type: Cost/Resource Use

Level of Analysis: Clinician Group

Data Source(s): Administrative Data (non-claims); Claims Data

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Rheumatoid Arthritis episode-based cost measure to the MIPS measure set. This measure evaluates a clinician's or clinician group's risk-adjusted and specialty-adjusted cost to Medicare for the management and treatment of rheumatoid arthritis. Clinicians have opportunities to improve their care for patients with rheumatoid arthritis and reduce costs, including diagnosing earlier, optimizing the use of monitoring tests, corticosteroids, and disease-modifying anti-rheumatic drugs (DMARDs), as well as improving levels of medication non-adherence. The Rheumatoid Arthritis measure will fill a measurement gap area in care for rheumatoid arthritis in the MIPS cost performance category, as there are currently no episode-based cost measures for this clinical area. CMS is considering this measure for use in MIPS because it has the potential to capture a large number of patients and Medicare spending and encourage performance improvement.

Table 3.9.1. MUC2023-209 Brief Summary of Measure Information

MERIT Submission Information MUC2023-209	Description
Measure name	Rheumatoid Arthritis
MUC ID	MUC2023-209
Cascade priority	Affordability and Efficiency
Measure steward	Centers for Medicare & Medicaid Services
Measure Developer	Acumen, LLC
Program submitted to	Merit-based Incentive Payment System- Cost
Committee assigned to	Clinician Committee

MERIT Submission Information MUC2023-209	Description
Related measures in the program	N/A
Is this a new measure in this year's MUC List?	Yes
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program	No
If previously used, please describe the history of the measure in CMS program	N/A
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	Yes
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
History of endorsement	N/A
Path to endorsement	Unknown
Measure Specification Details	
Measure Description	The Rheumatoid Arthritis 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 rheumatoid arthritis. This chronic condition measure includes the cost of services that are clinically related to the attributed clinician's role in managing care during a rheumatoid arthritis episode.
Data source	Administrative Data (non-claims); Claims Data
Level of analysis	Clinician: Group only
Numerator	The measure numerator is the weighted average ratio of the winsorized scaled standardized observed cost to the scaled expected cost for all rheumatoid arthritis episodes attributed to a clinician, where each ratio is weighted by each episode's number of days assigned to a clinician. This sum is then multiplied by the national average winsorized scaled observed episode cost to generate a dollar figure.
Denominator	The measure denominator is the total number of days from rheumatoid arthritis episodes assigned to the clinician across all patients.
Numerator exclusions	N/A
Denominator exclusions	The following standard exclusions are applied to ensure data completeness:

MERIT Submission Information MUC2023-209	Description
	• Patient has a primary payer other than Medicare for any time overlapping the episode window or 365-day lookback period prior to the episode window. • Patient was not enrolled in Medicare Parts A and B for the entirety of the lookback period plus episode window or was enrolled in Part C for any part of the lookback plus episode window. • Patient was not found in the Medicare Enrollment Database (EDB). • Patient's death date occurred before the episode end date. • Patient has an episode window shorter than 1 year. • Patients with extremely low treatment costs. • Patient's residence is outside the United States or its territories during the episode window, as indicated in the EDB.
Denominator exceptions	N/A
Risk adjustment	Yes
Development Status	Yes
If not fully developed, development stage	N/A
Target population	Medicare Fee for Service (FFS)
Measure type	Cost/Resource Use
Is the measure a composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.9.2. MUC2023-209 Measure Evaluation

MUC2023-209 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	Between 2012 and 2017, total Medicare spending for conventional disease-modifying anti-rheumatic drugs (cDMARDs) increased 5-fold, from $98 million to $579 million, while total Medicare spending on biologic disease-modifying anti-rheumatic drugs (bDMARDs) increased from $4.3 to $10.0 billion. Further, rheumatoid arthritis-related costs differ between effectively and non-effectively treated patients, with annual costs of $22,123 for the former and $9,250 for the latter. These variations in costs were attributable to more inpatient admissions, emergency department visits, and the number of prescription fills.[51]	--	The study population is the same as the target quality program population.
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	Standard exclusions are used to ensure data completeness. There are no exclusions specific to the Rheumatoid Arthritis cost measure. Although, standard exclusions are applied to ensure that the reportable episode populations are more homogenous and comparable than all episodes meeting the triggering logic for the measure.	The measure developer did not submit empirical evidence of conformance.	Most persons and entities in the quality program population are included in the specification.

[51] Stolshek BS, Wade S, Mutebi A, De AP, Wade RL, Yeaw J. Two-year adherence and costs for biologic therapy for rheumatoid arthritis. Am J Manag Care. 2018;24(8 Spec No.):SP315-SP321.

MUC2023-209 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	This is a claims-based measure. It can be calculated from claims and other data submitted electronically for billing and other purposes. All data elements are in defined fields in electronic sources.	No explicit articulation of people, processes, or technology required for measurement was submitted.	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden to data collection and reporting? **(***Context of Use)*	Data submitted by the measure developer show variation in the measure score at both TIN and TIN-NPI levels. At the TIN level, the mean score is $12,213.50, with a standard deviation of $4,386.01. The 10th percentile score is $7,432.84, the 50th percentile score is $11,729.81, and the 90th percentile score is $17,099.75.	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound? *(Context of Use)*	Reliability testing of the Rheumatoid Arthritis measure is conducted for individual clinicians (TIN-NPIs) and constructed using episodes ending between January 1, 2022, and December 31, 2022. At the 20-episode volume threshold, the testing results show that the mean reliability of the Rheumatoid Arthritis measure is 0.757 at the TIN-NPI level. With a minimum of 30 episodes per entity, the median reliability is 0.814. About 95% of the entities may have reliability above 0.6.	With a threshold of at least 10 episodes per entity, the median reliability is 0.657. The reliability for nearly 50% of the entities is below 0.6. When the threshold is increased to a minimum of 20 episodes per entity, the median reliability of is 0.77. The reliability for about 15% of the entities is below 0.6.	Some entities have reliability above the threshold (0.60) within a population that can be extrapolated to the program population.

MUC2023-209 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	Increased use of monitoring tests could lower healthcare utilization and related costs. Given the significant portion of costs related to pharmacotherapy for rheumatoid arthritis, this could improve patient care while minimizing the expensive use of less effective pharmaceutical treatments and therapies.[52] The existing literature shows that increasing DMARD use among patients with rheumatoid arthritis can improve patient outcomes and reduce costs.3 Although the ACR recommends the use of DMARDs in rheumatoid arthritis, only 36.2% of methotrexate (MTX) monotherapy users and 39.6% of multiple non-biologic DMARD users were found to receive care consistent with these recommendations after one physician visit; the percentages increased after two visits to 78.3% and 76.2%, respectively.[53] Data submitted by the measure developer show that, adverse events, Major procedures, outpatient evaluation and management (E&M) services, medications from parts B	The materials submitted did not include an explicit articulation of the way an entity may improve performance on the measure focus.	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population.

[52] Meyer R, Ellis LA, Bolge SC, Tkacz J, Kardel P, Reutsch C. National Quality Forum measures among rheumatoid arthritis patients in a large managed care population [Abstract]. 2013 ACR/ARHP Annual Meeting.
[53] Harrold LR, Reed GW, Kremer JM, et al. Identifying factors associated with concordance with the American College of Rheumatology rheumatoid arthritis treatment recommendations. Arthritis Res Ther. 2016;18:94. Published 2016 Apr 26. doi:10.1186/s13075-016-0992-3

MUC2023-209 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	and D, and outpatient physical, occupational, or speech and language pathology therapy are associated with worse measure scores. At the TIN-level, major procedures and part B and D drugs are associated with lower costs of adverse events. Outpatient E&M services and outpatient physical, occupational, or speech and language pathology therapy are associated with higher costs of adverse events, which may reflect higher clinical needs related to adverse events. This suggests these costs could be reduced without increasing adverse events. Data submitted by the measure developer show that, overall, the cost measure reflects both the cost directly related to treatment choices and the cost of related adverse outcomes. Therefore, there is evidence that the measure captures what it purports to measure. Adverse event costs are associated with a worse measure score. Major procedures, outpatient E&M services, medications from parts B and D, and outpatient physical, occupational, or speech and language pathology therapy are also associated with a worse measure score. However, major procedures and part B and D drugs are associated with lower costs of adverse events at the TIN reporting level, and the association is		

MUC2023-209 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	less clear at the TIN-NPI reporting level, consistent with these services being essential for patient outcomes. Outpatient evaluation and management services and outpatient physical, occupational, or speech and language pathology therapy are associated with higher costs of adverse events, which may reflect higher clinical needs related to adverse events.		
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	The measure developer notes they received generalized feedback on risk adjustment in episode-based cost measure calculation during a previous TEP meeting. The Clinician Expert Workgroup's feedback on risk adjustors and exclusions was incorporated into the model. The draft measure also underwent a national field-testing period and public comment periods, after which the Clinical Expert Workgroup further refined measure specifications based on feedback collected during testing. The risk adjustment model for this measure uses a linear regression model, which utilizes variables from the CMS Hierarchical Condition Code Version 24 (CMS-HCC V24) 2021 Risk Adjustment Model. This includes comorbidities captured by 86 HCC codes that map with thousands of ICD-10-CM codes, interaction	The materials submitted do not provide a detailed rationale for confounders included in the model.	N/A

MUC2023-209 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	variables accounting for a range of comorbidities, and other standard risk adjustors, including patient level demographics (i.e., age) and health status (i.e., disability status, end-stage renal disease [ESRD] status, recent use of long-term care).		
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	--	The submitted materials do not include an explicit articulation of the resources and context that might facilitate or be a barrier to the way an entity may improve.	Unable to determine if there is an articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-209 Measure Reliability[54]

The performance score is a risk- and specialty-adjusted average cost per episode.

Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across all entities. σ^2_{within} is the standard deviation of the score for a single entity.

For TIN-level entities with at least 10 episodes, the measure report indicates a median signal-to-noise reliability of 0.629 across 5,185 entities.

[54] Developer provided additional decile tables to replace simulated reliability calculations.

For TIN-level entities with at least 20 episodes, the measure report indicates a median signal-to-noise reliability of 0.770 across 3,051 entities.

For TIN-level entities with at least 30 episodes, the measure report indicates a median signal-to-noise reliability of 0.814 across 2,364 entities.

Decile tables:

The developer provided decile tables of performance score for the 20-episode threshold level and decile tables for reliability for the 10-, 20-, and 30-episode threshold levels.

For Tables 3.9.3, entities are sorted by performance score, and the average score by decile is reported along with the number of entities and episodes included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.9.3. Performance Score Deciles – TINs with at least 20 attributed episodes

MUC2023-209	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	$12,214 ($4,386)	$1,720	$7,433	$8,967	$9,974	$10,882	$11,730	$12,726	$13,759	$15,025	$17,100	$26,515	$53,193
Entities	3,051	1	305	305	305	305	305	305	305	305	305	305	1
# Episodes	484,635	--	--	--	--	--	--	--	--	--	--	--	--

For Tables 3.9.4, 3.9.5, and 3.9.8, entities are sorted by reliability and the average reliability by decile is reported. Average and standard deviation are also included.

Table 3.9.4. MUC2023-209 Mean Reliability (by Reliability Decile) – TINs with at least 10 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.63	0.23	0.300	0.409	0.500	0.581	0.657	0.725	0.795	0.855	0.914	0.974

Table 3.9.5. MUC2023-209 Mean Reliability (by Reliability Decile) – TINs with at least 20 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.74	0.18	0.489	0.582	0.658	0.719	0.770	0.816	0.86	0.896	0.935	0.950

Table 3.9.6. MUC2023-209 Mean Reliability (by Reliability Decile) – TINs with at least 30 attributed episodes

Mean	SD	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10
0.78	0.15	0.577	0.664	0.725	0.775	0.814	0.846	0.883	0.912	0.944	0.977

Interpretation:

With a threshold of at least 10 episodes per entity, the median reliability is 0.629. Nearly 50% of the entities may have reliability below 0.6.

When the threshold is increased to a minimum of 20 episodes per entity, the median reliability is 0.77. About 15% of the entities may have reliability below 0.6.When the threshold is increased to a minimum of 30 episodes per entity, the median reliability is 0.814. About 5% of the entities may have reliability below 0.6.

Merit-based Incentive Payment System-Quality Program

3.10 MUC2023-141 Positive PD-L1 Biomarker Expression Test Result Prior to First-Line Immune Checkpoint Inhibitor Therapy

Description: Percentage of patients aged 18 years and older, with a diagnosis of metastatic non-small cell lung cancer or squamous cell carcinoma of head and neck on first-line immune checkpoint inhibitor (ICI) therapy, who had a positive PD-L1 biomarker expression test result prior to giving ICI therapy.

Measure Type: Process

Level of Analysis: Clinician: Individual only

Data Source(s): Registries

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Positive PD-L1 Biomarker Expression Test Result Prior to First-Line Immune Checkpoint Inhibitor Therapy measure to the MIPS quality measure set. This fully developed process measure will address timely biomarker testing for patients with a diagnosis of metastatic non-small cell lung cancer or squamous cell carcinoma that impacts treatment decisions and improves patient outcomes. Appropriate and timely intervention of PD-L1 biomarker expression testing prior to initiation of first-line treatment for the metastatic non-small cell lung cancer or squamous cell carcinoma of head and neck can lead to improvements in mortality and morbidity and reduce healthcare cost by avoiding treatment delays and prescription of ineffective therapies. This measure fills a gap in MIPS for treatment of patients with lung cancer and may be a potential future addition to the Advancing Cancer Care MVP. This quality measure will support and possibly incentivize efforts to implement these necessary improvements to practice quality in the field of immunotherapy. While the current MIPS Oncology specialty measure set includes 25 measures, only 11 are specialty specific.

Table 3.10.1. MUC2023-141 Brief Summary of Measure Information

MERIT Submission Info MUC2023-141	Description
Measure name	Positive PD-L1 Biomarker Expression Test Result Prior to First-Line Immune Checkpoint Inhibitor Therapy
MUC ID	MUC2023-141
Cascade priority	Chronic Conditions
Measure steward	Society for Immunotherapy of Cancer (SITC)
Measure Developer	Society for Immunotherapy of Cancer (SITC)
Program submitted to	Merit-based Incentive Payment System-Quality
Committee assigned to	Clinician Committee
Related measures in the program	Similar measures include: 1) PIMSH8 (Practice Insights by McKesson in Collaboration with The US Oncology Network) (QCDR)Measure Title: Oncology: Mutation Testing for Lung Cancer Completed Prior to Start of Targeted Therapy Program: MIPS QCDR2) CAP34 (Pathologists Quality Registry) (QCDR)Measure Title: Biomarker Status to Inform Clinical Management and Treatment Decisions in Patients with Non-small Cell Lung Cancer Program: MIPS QCDR3) NPQR15 (National Pathology Quality Registry) (QCDR)Measure Title: Non-small cell lung carcinoma (NSCLC) ancillary biomarker testing status and turnaround time (TAT) from point of specimen accession date to ancillary biomarker testing completion and reporting date should be = 10 days Program: MIPS QCDR4) CBE 1859/QPP 451 (American Society of Clinical Oncology- steward) Measure Title: RAS (KRAS and NRAS) Gene Mutation Testing Performed for Patients with Metastatic Colorectal Cancer who receive Anti-epidermal Growth Factor Receptor (EGFR) Monoclonal Antibody Therapy Program: MIPS Program Data Source: CQM5) CBE 1860/QPP 452 (American Society of Clinical Oncology- steward) Measure Title: Patients with Metastatic Colorectal Cancer and RAS (KRAS or NRAS) Gene Mutation Spared Treatment with Anti-epidermal Growth Factor Receptor (EGFR) Monoclonal Antibodies Program: MIPS Program Data Source: CQM
Is this a new measure in this year's MUC List?	No
If not a new measure, then describe the history of this measure in prior MUC list inclusion	Submitted previously but not included in MUC List
Is the measure currently used in a CMS program	N/A

MERIT Submission Info MUC2023-141	Description
If previously used, please describe the history of the measure in CMS program	N/A
Any other program the measure is in use	No
Is this measure being proposed to meet a statutory requirement?	N/A
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
History of endorsement	N/A
Path to endorsement	Year of next anticipated CDP endorsement review: 2024
Measure Specification Details	
Measure Description	Percentage of patients, aged 18 years and older, with a diagnosis of metastatic non-small cell lung cancer or squamous cell carcinoma of head and neck on first-line immune checkpoint inhibitor (ICI) therapy, who had a positive PD-L1 biomarker expression test result prior to giving ICI therapy.
Data source	Registries
Level of analysis	Clinician: Individual only
Numerator	Patients who had a positive PD-L1 biomarker expression test result prior to the initiation of first-line immune checkpoint inhibitor therapy. Numerator Guidance: • PD-L1 biomarker expression test – FDA approved test that measures the expression of PD-L1 on cancer and/or immune cells. • Positive PD-L1 biomarker expression test result – PD-L1 test is considered positive if the cancer and/or immune cells have an appropriate threshold of PD-L1 expression based on the approved companion diagnostic.
Denominator	Patients, 18 years and older, with a diagnosis of metastatic non-small cell lung cancer or squamous cell carcinoma of head and neck and on first-line immune checkpoint inhibitors without chemotherapy. Denominator Guidance: • Immune checkpoint inhibitors-class of medications that prevent tumors from “hiding” or “evading” the body’s natural immune system. This is a form of cancer immunotherapy. Immune checkpoint inhibitor medications

MERIT Submission Info MUC2023-141	Description
	include PD-1 inhibitor drugs, PD-L1 inhibitor drugs, and CTLA-4 inhibitor drugs. • PD-1 inhibitors drugs include: Pembrolizumab, Nivolumab, Cemiplimab • PD-L1 inhibitor drugs include: Atezolizumab • CTLA-4 inhibitor drugs include Ipilimumab • First-line treatment- initial, or first treatment recommended for cancer • Various treatment regimens were considered, including immune checkpoint inhibitors. • PD-L1 testing required per FDA approval for the applicable histology • If the patient is on any of the below immune checkpoint inhibitor(s) as first-line treatment for metastatic disease, they must also have one of the specific subsets of non-small cell lung cancer (NSCLC) or squamous cell carcinoma of head and neck (HNSCC) ○ Pembrolizumab (PD-1 inhibitor drug) AND first-line treatment in patients with metastatic NSCLC OR first-line treatment in patients with metastatic squamous cell carcinoma of the head and neck OR ○ Cemiplimab (PD-1 inhibitor drug) AND first-line treatment in patients with metastatic NSCLC OR ○ Atezolizumab (PD-L1 inhibitor drug) AND first-line treatment in patients with metastatic NSCLC OR ○ Nivolumab (PD-1 inhibitor drug) and Ipilimumab (CTLA-4 inhibitor drug) combination AND first-line treatment in patients with metastatic NSCLC.
Numerator exclusions	N/A
Denominator exclusions	Patients with metastatic non-small cell lung cancer (NSCLC) with EGFR mutations, ALK translocations, or other targetable genomic abnormalities with approved first-line targeted therapy, such as NSCLC with ROS1 rearrangement, BRAF V600E mutation, NTRK 1/2/3 gene fusion, MET ex14 skipping mutation, and RET rearrangement.
Denominator exceptions	Documentation of medical reason(s) for not performing the PD-L1 biomarker expression test prior to initiation of first-line immune checkpoint inhibitor therapy (e.g., patient is in an urgent or emergent situation where delay to treatment would jeopardize the patient's health status; other medical reasons/contraindication). Denominator Exceptions Guidance:

MERIT Submission Info MUC2023-141	Description
	• PD-L1 biomarker expression testing was unable to be performed prior to the initiation of first-line immune checkpoint inhibitor therapy due to an urgent or emergent situation where any treatment delay would jeopardize the patient's health and/or cancer care. • Lack of available tissue for PD-L1 biomarker expression testing due to a documented medical and/or surgical contraindication which would not allow for the patient to undergo a tissue biopsy safely.
Risk adjustment	Yes
Development Status	Fully Developed
If not fully developed, development stage	N/A
Target population	Medicare Fee for Service
Measure type	Process
Is the measure composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.10.2. MUC2023-141 Measure Evaluation

MUC2023-141 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Importance:	Timely and appropriate expression of the PD-L1 biomarker prior to initiation of first-line treatment for the metastatic	--	The study population is the same as the target quality program population.

MUC2023-141 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Does the measure align with goals and priorities? (*Concept of Interest)*	non-small cell lung cancer or squamous cell carcinoma of head and neck can lead to improvements in mortality and morbidity. Without this testing, appropriate treatments could be delayed, or inappropriate treatments prescribed, resulting in worse outcomes for patients and higher costs.[55, 56] Results from the MYLUNG Consortium research protocol, which retrospectively reviewed the charts of 3,474 adult patients with non-small cell lung cancer from over 1,000 providers found that less than 50% of patients were tested for five major biomarkers (ALK, BRAF, EGFR, PD-L1, and ROS-1).[57] This measure is intended to enhance compliance with the clinical guidelines by ensuring eligible providers are completing timely biomarker testing to support treatment decisions.		

[55] Pai, S., Blaisdell, D., Brodie, R., et al. (2020). Defining current gaps in quality measures for cancer immunotherapy: consensus report from the Society for Immunotherapy of Cancer (SITC) 2019 Quality Summit. Journal for ImmunoTherapy of Cancer, 8, e000112. doi:10.1136/jitc-2019-000112. https://jitc.bmj.com/content/8/1/e000112

[56] Lim, C., Tsao, M.S., Le, L.W., et al. (2015). Biomarker testing and time to treatment decision in patients with advanced nonsmall-cell lung cancer. Annals of Oncology, 26, 1415–1421. https://www.sciencedirect.com/science/article/pii/S092375341934517X?via%3Dihub

[57] Robert, N.J., Nwokeji, E., Espirito, J.L., et al. (2021). Biomarker tissue journey among patients with untreated metastatic non-small cell lung cancer (mNSCLC) in the U.S. Oncology Network community practice. 2021 ASCO Annual Meeting. Abstract 9004. Presented June 4, 2021.

MUC2023-141 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	This measure includes all patients aged 18 years and older, with a diagnosis of metastatic non-small cell lung cancer or squamous cell carcinoma of the head and neck on first-line immune checkpoint inhibitor (ICI) therapy without chemotherapy. At the program level, this measure was determined to be implementable within MIPS and the exclusion/exceptions represent the typical analytic found within MIPS quality measures. At the clinical level, the measure has evidence based on the clinical guidelines and data element testing was 'sound' for the exclusion/exception data elements.	--	Most persons and entities in the quality program population are included in the specification.
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	The measure is based on evidence-based clinical guidelines. This is a registry-based measure. The developer notes that some data elements are in defined fields in electronic sources. The measure has been determined to be feasible for implementation in the MIPS program as specified.	The measure developer did not submit empirical evidence demonstrating collection or public reporting of this measure leads to positive or intended consequences. Further, the measure developer did not submit an explicit articulation of the people, processes, and technology required for data collection and reporting.	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.

MUC2023-141 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden to data collection and reporting? (*Context of Use)*	Two clinical guidelines recommend positive PD-L1 biomarker expression test prior to giving first-line immune checkpoint inhibitor therapy in the metastatic non-small cell lung cancer or squamous cell carcinoma of head and neck population.[58, 59] The developer notes that some data elements are available in defined fields. However, no information was submitted describing the burden of data collection and reporting. Pilot testing conducted by the measure developer found variation in measure performance across measured entities.	The measure developer notes that some data elements are available in defined fields. However, no information was submitted describing the burden of data collection and reporting.	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound? *(Context of Use)*	The measure developer conducted signal-to-noise analysis with a sample of 51 clinicians. The reliability of the measure scores ranged from 0.695 to 1.00. The median reliability score was 0.947 and the average reliability score was 0.918, with a 95% confidence interval from 0.887 to 0.949. Overall, 92% of clinicians had measure scores with reliabilities of 0.70 or higher.	--	Most or all entities have reliability above the threshold (0.60) within a population that extrapolates to the quality program population.
Validity: May providers/facilities/care	In a survey of 27 subject matter experts conducted by the measure developer, 74% (20 of 27) agreed that	In a survey of 27 subject matter experts conducted by the measure developer, seven (26%)	There is an association between the entity and the measure focus in

[58] National Comprehensive Cancer Network (2021). NCCN Clinical Practice Guidelines in Oncology: Head and Neck Cancer. https://www.nccn.org/professionals/physician_gls/pdf/head-and-neck.pdf
[59] National Comprehensive Cancer Network (2021). NCCN Clinical Practice Guidelines in Oncology: Non-Small Cell Lung Cancer. https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf

MUC2023-141 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
systems effectively improve on this measure? *(Context of Use)*	the measure could effectively distinguish between good and poor quality of care provided by individual clinicians. Out of these 20 experts, 4 strongly agreed and 16 agreed with the measure's ability to distinguish between good and poor-quality care.	were neutral or disagreed that the measure could effectively distinguish between good and poor quality of care. These experts advised considering both immunotherapy and chemotherapy.	a population that extrapolates to the quality program population.
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	This measure is not risk adjusted.	--	N/A
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	This clinical process measure would provide data regarding the performance of the action by the clinician. This data would then support changes to the clinicians' workflow that facilitate changes to improve the patients' outcomes.		There is not an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-141 Simulated Measure Reliability Tables

The performance score is a percentage of patients with a positive test result for each entity. Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across the entities. σ^2_{within} is the variance (standard deviation squared) of the score within a single entity. The measure report indicates a median signal-to-noise reliability of 0.95.

Simulated decile tables:

Computer simulation was used to create a dataset that mirrors, as closely as possible, the mean, standard deviation, and percentile information provided for the performance score and calculated reliability. Tables 3.10.3 and 3.10.4 are created from the simulated

dataset and provide reviewers with a more standardized format to assess reliability. For Table 3.10.3, entities were sorted by performance score, and the average score by decile (estimated from the simulated data) is reported along with the number of entities included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.10.3. MUC2023-141 Importance (Decile by performance score)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	19.29 (32.53)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.25	26.49	63.14	96.04	100.00
Entities	15000	9827	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	1134

For Table 3.10.4, entities were sorted by reliability and the average reliability by decile (estimated from the simulated data) is reported along with the number of entities included in each average. Average, minimum, and maximum reliability and expected events are also included.

Table 3.10.4. MUC2023-141 Reliability (Decile by reliability)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Reliability	0.94	0.70	0.71	0.78	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Entities	15000	601	1500	1500	1500	1500	1500	1500	1500	1500	1500	1500	10961

Assumptions:

The study includes 51 entities but does not report the total number of patients included. The measure report estimates an annual denominator size of 93,555, or that many total patients. A 2023 report by Definitive Healthcare estimates there are about 15,000 oncologists in the United States.

Interpretation:

The median reported reliability is 0.95. The simulated data suggest that 100% of the entities may have a reliability greater than 0.6.

3.11 MUC2023-161 Appropriate Germline Testing for Ovarian Cancer Patients

Description: Percentage of patients, aged 18 and older, diagnosed with epithelial ovarian, fallopian tube, or primary peritoneal cancer who undergo germline testing within 6 months of diagnosis.

Measure Type: Process

Level of Analysis: Clinician: Individual and Group

Data Source(s): Electronic Health Record; Registries

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Appropriate Germline Testing for Ovarian Cancer Patients measure to the MIPS quality measure set. This fully developed process measure is a new measure for oncology that addresses patients diagnosed with epithelial ovarian, fallopian tube, or primary peritoneal cancer who undergo germline testing within 6 months of their diagnosis. Knowledge about underlying molecular alterations in ovarian cancer could allow for more personalized diagnostic, predictive, prognostic, and therapeutic strategies for the patient. This measure fills a gap in MIPS as it follows the American Society of Clinical Oncology guidelines for testing for germline BRCA mutations, to facilitate better guided treatment for epithelial ovarian, fallopian tube or primary peritoneal cancer for improved outcomes. Additionally, this genetic testing would help inform patient risk of other cancers and the potential need for cascade testing of family members. Additionally, this measure fills a current quality measure inventory gap within the oncologic clinical topic and is a concept not duplicative of quality measures currently in MIPS. While the Oncology specialty measure set includes 25 measures, only 11 are specialty specific. This measure could also be considered for inclusion within the Advancing Cancer Care MVP.

Table 3.11.1 MUC2023-161 Brief Summary of Measure Information

MERIT Submission Information MUC2023-161	Description
Measure name	Appropriate Germline Testing for Ovarian Cancer Patients
MUC ID	MUC2023-161
Cascade priority	Chronic Conditions
Measure steward	American Society of Clinical Oncology

MERIT Submission Information MUC2023-161	Description
Measure Developer	American Society of Clinical Oncology
Program submitted to	Merit-based Incentive Payment System-Quality
Committee assigned to	Clinician Committee
Related measures in the program	CMIT Measure ID: 1659 Mismatch Repair (MMR) or Microsatellite Instability (MSI) Biomarker Testing Status in Colorectal Carcinoma, Endometrial, Gastroesophageal, or Small Bowel Carcinoma.
Is this a new measure in this year's MUC List?	Yes
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program	N/A
If previously used, please describe the history of the measure in CMS program	N/A
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	N/A
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
History of endorsement	N/A
Path to endorsement	Anticipated CDP endorsement review in 2024
Measure Specification Details	
Measure Description	Percentage of patients, aged 18 and older, diagnosed with epithelial ovarian, fallopian tube, or primary peritoneal cancer who undergo germline testing within 6 months of diagnosis.
Data source	Electronic Health Record; Registries
Level of analysis	Clinician: Individual and Group (MIPS-Quality only)
Numerator	Patients who receive germline genetic testing for BRCA1 and BRCA2 (ideally within the context of a multigene panel) or who have genetic counseling completed within 6 months of diagnosis. Numerator guidance: The ASCO guideline panel recommends that germline sequencing of BRCA1 and BRCA2 be performed in the context of a multigene panel that includes BRCA1, BRCA2, RAD51C, RAD51D, BRIP1, MLH1, MSH2, MSH6, PMS2, and PALB2. While the TEP prefers germline genetic testing is conducted for other ovarian cancer susceptibility genes in

MERIT Submission Information MUC2023-161	Description
	addition to BRCA1 and BRCA2 as recommended in the guideline, this measure focuses specifically on BRCA1 and BRCA2 as there may be payer variation or other limitations in the availability of multigene panels. While the ASCO guideline recommendation calls for germline testing to be conducted at the time of diagnosis, the TEP chose to specify the time period for germline testing to occur within 6 months after diagnosis. The performed/collected date for BRCA testing will be used to calculate the numerator time period.
Denominator	All patients, aged 18 and older, with epithelial ovarian, fallopian tube, or primary peritoneal cancer newly diagnosed between July of the previous calendar year through June of the measurement period with two encounters during the measurement period
Numerator exclusions	N/A
Denominator exclusions	Patients who have germline BRCA testing completed before diagnosis of epithelial ovarian, fallopian tube, or primary peritoneal cancer. Denominator exclusion guidance: The denominator further specifies a population of patients with unknown BRCA status. The panel acknowledges patients diagnosed with ovarian cancer may have undergone germline testing for a previous breast cancer diagnosis or that patients may have ovarian cancer identified at the time of prophylactic surgery following germline testing; patients with a known BRCA status at the time of ovarian cancer diagnosis are therefore not a population for whom this measure applies.
Denominator exceptions	None. Denominator exception guidance: The measure was originally specified and tested with a denominator exception for documentation of patient decline or refusal of testing. Feasibility and data element validity testing indicates that the exception is not justified given the empirical evidence.
Risk adjustment	No
Development Status	Fully Developed
If not fully developed, development stage	N/A
Target population	Medicare Fee for Service, All Payer
Measure type	Process
Is the measure a composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No

MERIT Submission Information MUC2023-161	Description
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.11.2. MUC2023-161 Measure Evaluation

MUC2023-161 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	A nationwide study involving ovarian and select breast cancer patients. Participants were identified in the Flatiron Health database and were required to have demonstrated recent and persistent underutilization of germline genetic testing.[60] Specifically, within one year of diagnosis, 60% of those with ovarian cancer lacked documentation of germline testing. Further, several significant disparities related to age, race, and insurance status were identified. Another study submitted as evidence noted that despite recommendations for women who have been diagnosed with ovarian cancer to undergo genetic testing,	--	While the study population differs from the target quality program population, the importance for the selected program population can be extrapolated.

[60] Lau-Min, K. S., McCarthy, A. M., Nathanson, K. L., & Domchek, S. M. (2023). Nationwide Trends and Determinants of Germline BRCA1/2 Testing in Patients With Breast and Ovarian Cancer, Journal of the National Comprehensive Cancer Network, 21(4), 351-358.e4. Retrieved Apr 21, 2023, from https://jnccn.org/view/journals/jnccn/21/4/article-p351.xml

MUC2023-161 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	merely 30% underwent any genetic testing.[61]		
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	Standard exclusions are used to ensure data completeness. Entities in the measure focus and target population are included in the specification.	--	Most persons and entities in the quality program population are included in the specification.
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	Measure's use of EHR & registry data as well as option for telehealth visits included in specification lowers burdens to data collection.	--	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden to data collection and reporting? **(***Context of Use)*	Expected outcomes include personalized/targeted treatment recommendations related to mutation status (i.e., improved patient outcomes), as well as informed patient risk of other cancers and identification of the need for cascade testing of family members (screening and prevention).	--	All the performance improvements to the benchmark have a significant impact on quality program population outcomes.

[61] Konstantinopoulos, P. A., Norquist, B., Lacchetti, C., Armstrong, D., Grisham, R. N., Goodfellow, P. J., Kohn, E. C., Levine, D. A., Liu, J. F., Lu, K. H., Sparacio, D., & Annunziata, C. M. (2020). Germline and Somatic Tumor Testing in Epithelial Ovarian Cancer: ASCO Guideline. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 38(11), 1222–1245. https://doi.org/10.1200/JCO.19.02960

MUC2023-161 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Reliability: Is measure performance scientifically sound? *(Context of Use)*	Signal-to-noise reliability testing was performed at the Clinician – Group level of analysis, with a sample size of 12 practices. The median reliability across clinicians is about 0.89, and the median reliability across groups is about 0.86. None of the clinicians or the groups have reliability below 0.6.	--	Most or all entities have reliability above the threshold (0.60) within a population that extrapolates to the quality program population.
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	Data element validity testing results indicated the measure's data elements are feasible and valid, with 52.3 percent mean performance score among the 32 individual clinicians analyzed in testing.	--	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population.
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	This measure is not risk adjusted.	--	N/A
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	This measure will be collected as a Clinical Quality Measure (CQM) Registry measure. Once more performance data is obtained, the benchmark for this measure will be established.	--	There is an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-161 Measure Reliability[62]

The performance score is the percentage of patients that undergo testing for each entity.

Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across the 12 entities. σ^2_{within} is the variance (standard deviation squared) of the score within a single entity. Among the 12 entities included, there was an average of 24.7 patients per entity with a minimum of 12 patients needed for an entity to qualify for testing.

Supplemental information provided by the developer indicates a range of signal-to-noise reliability from 0.684 to 1.0 across 32 clinicians, and a range of signal-to-noise reliability from 0.770 to 0.988 across 12 groups.

Decile tables:

The developer provided decile tables of performance score and reliability by clinician and by group.

For Tables 3.11.3 and 3.11.4, entities are sorted by performance score, and the average score by decile is reported along with the number of entities included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.11.3. MUC2023-161 Importance (Decile by performance score) - by Clinician

MUC2023-161	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	0.52	0.09	0.11	0.12	0.15	0.46	0.52	0.67	0.78	0.88	1.00	1.00	1.00
Entities	32	1	3	3	3	3	3	4	3	3	3	4	4
Total Persons	296	11	29	27	26	26	28	37	29	30	26	38	38

[62] Developer provided additional decile tables to replace simulated reliability calculations.

Table 3.11.4. MUC2023-161 Importance (Decile by performance score) - by Group

MUC2023-161	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	0.69	0.12	0.20	0.67	0.71	0.74	0.77	0.79	0.80	0.84	0.94	0.96	0.96
Entities	12	1	2	1	1	1	1	1	2	1	1	1	1
Total Persons	296	57	119	17	12	15	13	14	35	12	12	47	47

For Tables 3.11.5 and 3.11.6, entities are sorted by reliability and the average reliability by decile is reported along with the number of entities included in each average, and the average number of expected events. Average, minimum, and maximum reliability and expected events are also included.

Table 3.11.5. MUC2023-161 Reliability (Decile by reliability) - by Clinician

MUC2023-161	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Reliability	0.866	0.684	0.724	0.785	0.825	0.850	0.880	0.892	0.908	0.936	0.997	1.000	1.000
Entities	32	1	4	4	3	2	3	4	2	3	3	4	4
Total Persons	296	8	38	37	29	17	26	36	18	27	30	38	38

Table 3.11.6. MUC2023-161 Reliability (Decile by reliability) - by Group

MUC2023-161	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Reliability	0.892	0.770	0.805	0.818	0.823	0.840	0.855	0.866	0.918	0.961	0.970	0.988	0.988
Entities	12	1	2	1	1	1	1	1	1	1	2	1	1
Total Persons	296	12	25	15	17	14	12	17	18	12	119	47	47

Interpretation:

The median reliability across clinicians is about 0.89, and the median reliability across groups is about 0.86. None of the clinicians or the groups have reliability below 0.6.

3.12 MUC2023-162 Patient-Reported Pain Interference Following Chemotherapy among Adults with Breast Cancer

Description: The PRO-PM will assess pain interference following chemotherapy administered with curative intent to adult patients with breast cancer.

Measure Type: PRO-PM or Patient Experience of Care

Level of Analysis: Clinician: Individual and Group

Data Source(s): Electronic Health Record; Paper Medical Records; PROMIS Scale: Pain Interference Short Form 4a; Patient Reported Data and Surveys

Development Status: Fully Developed

Endorsement Status: Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Patient-Reported Pain Interference Following Chemotherapy among Adults with Breast Cancer measure to the MIPS quality measure set. This patient-reported outcome-based performance quality measure would fill a gap by providing the patient's experience of care related to breakthrough pain after chemotherapy for breast cancer. Additionally, as a patient reported outcome measures, it fills a current CMS high priority measure inventory gap within the oncologic clinical topic and is a concept not duplicative of quality measures currently in MIPS. Data from this measure provides insight into the effectiveness of medical oncologists in helping patients to minimize the persistent impact of their treatments and is useful to inform practice improvement. While the Oncology specialty measure set includes 25 measures, only 11 are specialty specific. This would be the first outcome specialty specific oncology measure to address the patient experience of care. There is potential consideration for adding broader cancer diagnosis such as colon and lung cancer to this measure in the future.

Table 3.12.1. MUC2023-162 Brief Summary of Measure Information

MERIT Submission Information MUC2023-162	Description
Measure name	Patient-Reported Pain Interference Following Chemotherapy among Adults with Breast Cancer
MUC ID	MUC2023-162
Cascade priority	Person-Centered Care
Measure steward	Purchaser Business Group on Health
Measure Developer	Purchaser Business Group on Health

MERIT Submission Information MUC2023-162	Description
Program submitted to	Merit-based Incentive Payment System-Quality
Committee assigned to	Clinician Committee
Related measures in the program	N/A
Is this a new measure in this year's MUC List?	Yes
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program	N/A
If previously used, please describe the history of the measure in CMS program	N/A
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	N/A
CBE endorsement status	Endorsed
CBE endorsement number if applicable	CBE 3718
History of endorsement	Year of most recent endorsement: 2023
Path to endorsement	Year of next anticipated CDP endorsement review: 2026
Measure Specification Details	
Measure Description	The PRO-PM will assess pain interference following chemotherapy administered with curative intent to adult patients with breast cancer.
Data source	Electronic Health Record; Paper Medical Records; PROMIS Scale: Pain Interference Short Form 4a; Patient Reported Data and Surveys
Level of analysis	Clinician: Individual and Group (MIPS-Quality only)
Numerator	The PRO-PM numerator is the mean of the patient-level PROMIS Pain Interference scores at the follow-up survey.
Denominator	Adult patients with stages I-III female breast cancer receiving an initial chemotherapy regimen. Denominator details: The denominator population includes the following patients: ≥ age 18 on the date of diagnosis, AND Stages I-III female breast cancer AND Receiving an initial chemotherapy regimen with a defined duration at the test site Patients with baseline and follow-up PROMIS surveys

MERIT Submission Information MUC2023-162	Description
	(See the MIF/Data Dictionary for additional definitions)
Numerator exclusions	N/A
Denominator exclusions	• Patients on a therapeutic clinical trial • Patients with recurrence/disease progression • Patients who leave the practice during the follow-up period • Patients who die during the follow-up period
Denominator exceptions	N/A
Risk adjustment	Yes
Development Status	Fully Developed
If not fully developed, development stage	N/A
Target population	Medicare Fee for Service, Medicare Advantage, Medicaid, All Payer. All adult cancer patients not restricted by payer type.
Measure type	PRO-PM or Patient Experience of Care
Is the measure composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.12.2. MUC2023-162 Measure Evaluation

MUC2023-162 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	Patient-reported outcome (PRO) assessment related to cancer care (e.g., symptom burden following chemotherapy), as well as development of PRO-based performance measures (PRO-PMs) has been endorsed widely.[63, 64] Oncologist assessment and management of symptoms during chemotherapy has been shown to benefit patients by lowering symptom burden, reducing suffering, and providing support into the survivorship phase. Evidence suggests that when oncologists provide high quality care to patients receiving curative-intent cytotoxic therapy, symptom burden can be reduced and patient transition into the cancer survivorship period can be improved.[65] [66]	--	The study population is the same as the target quality program population.

[63] Valderas, J.M., Kotzeva, A., Espallargues, M., et al. (2008). The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. Qual Life Res 17(2):179-193.
[64] Chen, J., Ou, L., & Hollis, S. J. (2013). A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Services Research, 13(1). https://doi.org/10.1186/1472-6963-13-211
[65] Bubis, L. D., Davis, L., Mahar, A., Barbera, L., Li, Q., Moody, L., Karanicolas, P., Sutradhar, R., & Coburn, N. G. (2018). Symptom burden in the first year after cancer diagnosis: An analysis of patient-reported outcomes. Journal of Clinical Oncology, 36(11), 1103-1111. https://doi.org/10.1200/jco.2017.76.0876
[66] National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology, Cancer-Related Fatigue, Version 2.2022. NCCN, 2022 https://www.nccn.org/guidelines/category_3

MUC2023-162 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	Specification notes sample of adults with stages I-III female breast cancer that are currently receiving an initial chemotherapy regimen. Specification includes appropriate entities for measure focus and target population.	--	Most persons and entities in the quality program population are included in the specification.
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	Some data elements are in defined fields in electronic sources. Burden for providers was assessed across 9 sites via a questionnaire. Respondents reported the majority of the implementation burden was associated with administering the survey rather than collecting the clinical and demographic data elements; patient identification was also a challenge, which test sites mitigated by building EHR reports to facilitate patient identification. Alpha testing conducted July 1, 2019, to September 5, 2019, calculated eligible data elements, missing responses, invalid response options, and inappropriate answers. Results informed the 1) revision of data elements related to skipped, not applicable, and missing variables, 2) development of a REDCap form for Beta testing, 3) training of project managers and abstractors, and 4) quality control procedures.	--	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.

MUC2023-162 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	Beta testing was conducted from October 1, 2019, to March 31, 2021. Testing involved 20 sites; however, due to limited patient responses, data from 10 sites were analyzed. Beta test results informed specifications (e.g., numerator, denominator, denominator exclusions, and risk adjustment variables/model).		
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden to data collection and reporting? (*Context of Use)*	Recent literature suggests cancer patient outcomes are improved when symptoms are reported during care (e.g., increased survival, reduced symptom burden and improved patient experience). Of particular interest is pain interference for adult patients with breast cancer after chemotherapy. Wiffen et al., (2017) reported moderate to severe pain as a common symptom for 30-50% of cancer patients (totaling approximately 510,000 to 850,000 annually).[67] Research also indicates that cancer survivors experience chronic pain which increases distress and negatively impacts quality of life. Findings show up to 40% of cancer survivors report chronic pain, and	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.

[67] Wiffen, P.J., Wee, B., Derry, S., et al. (2017). Opioids for cancer pain an overview of Cochrane reviews. Cochrane Database of Syst Rev 7:CD01292.

MUC2023-162 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	survivors with depression experience increased levels of pain.[68]		
Reliability: Is measure performance scientifically sound? *(Context of Use)*	Signal-to-noise reliability analysis was conducted (n=22). Median result was 0.39. Hierarchical linear regression modeling was used to evaluate measure reliability. Clinician-level reliability testing estimated the adjusted ICC was 0.08. The reliability estimate with the average sample size for a clinician (7.8 patients per clinician) was 0.42. The Spearman-Brown prophecy formula was employed and results indicated that an average sample size of 26 patient respondents was required to obtain a nominal reliability of 0.70. No clinicians in the sample had reliability at or above 0.70. Group-level reliability testing estimated the adjusted ICC was 0.097. The reliability estimate with the average sample size for a group (32 patients per group) was 0.77. The Spearman-Brown prophecy formula was used again, and results indicated an average sample size of 22 patient respondents would be required for sufficient reliability (i.e., at or above 0.70). Group specific reliability ranged	The reliability for groups with less than 16 patients may be below 0.6.	Some entities have reliability above the threshold (0.60) within a population that can be extrapolated to the program population.

[68] Glare, P.A., Davies, P.S., Finlay, E., Gulati, A. et al. (2014) Pain in Cancer Survivors. *Journal of Clinical Oncology, 32*(16): 1739-1747.

MUC2023-162 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	from 0.39 to 0.88 (M=0.66; SD=0.20), and a median reliability of 0.68. In the sample, 50% of groups had reliability scores at or above 0.70. The reliability for groups with at least 16 patients may be above 0.6.		
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	USPSTF Grade A, Strong recommendation or similar. Empiric validity was also tested (n=7), yielding a result of -0.567. A Face Validity Panel of 12 oncologists was convened in 2022. Seven of eight oncologists agreed that the measure was found to differentiate between good and poor quality (four declined to participate). Further, to test empiric validity, the measure developer collected data from test sites during the testing time period for HCAHPS, Outpatient Oncology Press Ganey, and QOPI. Without viewing submitted data, TEP members rated expected correlation strength between the PROMOnc measures and these available data. The measure developer then analyzed correlations for any measure for which the TEP hypothesized a moderate association and for which they had data for at least 7 test sites. The measure was moderately correlated with other patient-reported	--	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population.

MUC2023-162 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	outcome measures in the predicted direction, indicating the measure is associated with better reported patient experience.		
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	Risk adjustment variables were clearly defined and appropriate for the measure focus and target population.	--	N/A
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	There is an explicit articulation of the resources and context that might facilitate or be a barrier to the way an entity may improve.	--	There is an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-162 Measure Reliability

The performance score is the average patient-level score for each entity. Simulated reliability tables were developed given the reliability method chosen and availability of data.

Interpretation:

The reported reliability at the group level is 0.77 at the average sample size of 32 patients per group. Using the Spearman-Brown prophecy formula, the reliability for groups with less than 16 patients is below 0.6.

3.13 MUC2023-164 Adult COVID-19 Vaccination Status

Description: Percentage of patients aged 18 years and older seen for a visit during the performance period that are up to date on their COVID-19 vaccinations as defined by CDC guidelines on current vaccination

Measure Type: Process

Level of Analysis: Clinician: Individual and Group

Data Source(s): Registries

Development Status: Not Fully Developed, Specification

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Adult COVID-19 Vaccination Status measure to the MIPS quality measure set as this process measure represents a CMS high priority clinical topic due to the recent public health emergency and fulfills a gap in MIPS by addressing COVID-19 vaccination status. Widespread vaccination to prevent severe COVID-19 infection will be critically important to stemming the morbidity and mortality caused by this disease. The measure was updated for this MUC cycle to include the most recent CDC guidelines for the numerator. Currently, the MIPS quality measure set includes multiple quality measures that assess for vaccine administration that have been implemented for multiple years, however, such measures do not include COVID-19 vaccination. This measure, along with other activities, is part of larger federal effort to promote and track vaccine uptake. As vaccine uptake is partially driven by patients asking for the vaccine and clinicians administering it to eligible patients, the patient/clinician relationship is important for ensuring that patients are vaccinated. Therefore, this clinician-level measure will provide useful information about the success of vaccination efforts at the point of care and represent a priority topic to engage clinicians in quality improvements that drive health outcomes for their patients.

Table 3.13.1. MUC2023-164 Brief Summary of Measure Information

MERIT Submission Information MUC2023 164	Description
Measure name	Adult COVID-19 Vaccination Status
MUC ID	MUC2023-164
Cascade priority	Wellness and Prevention
Measure steward	Centers for Medicare & Medicaid Services

MERIT Submission Information MUC2023 164	Description
Measure Developer	National Committee for Quality Assurance
Program submitted to	Merit-based Incentive Payment System- Quality
Committee assigned to	Clinician Committee
Related measures in the program	COVID-19 Vaccination Coverage among Healthcare Personnel COVID-19 Vaccine: Percent of Patients/Residents Who Are Up to Date
Is this a new measure in this year's MUC List?	No
If not a new measure, then describe the history of this measure in prior MUC list inclusion	This measure was submitted as MUC20 -0045 COVID-19 Vaccination by Clinicians to the Merit-based Incentive Payment System-Quality Program and was reviewed by the MAP Rural Health Advisory Group, Clinician Workgroup, and the Coordinating Committee leading to a recommendation of conditional support. The MAP reviewed this measure in combination with several other de novo COVID-19 vaccination measures applicable to several settings of care. The MAP rationale was the same for all the COVID-19 vaccination measures: "MAP offered conditional support for rulemaking contingent on CMS bringing the measures back to the MAP once the specifications are further refined, CMS considering an expedited process for the measures for both NQF and CMS, and CMS exploring the inclusion of pediatric hospitals within the COVID measures." The last statement about the inclusion of pediatric hospitals is not relevant for this measure as this measure is focused on the ambulatory setting. The measure was also submitted as MUC2022-052 Adult COVID-19 Vaccination Status to the Merit-based Incentive Payment System-Quality Program and was reviewed by the MAP Rural Health Advisory Group, Clinician Workgroup, and the Coordinating Committee leading to a recommendation of "do not support." MAP discussed comments that there is regional variation in vaccine hesitancy and noted the measure does not have an exclusion for patient refusal. MAP also noted the measure does not include bivalent booster vaccinations and suggested that the developer consider future updates to the measure specification by defining vaccination as "up-to-date vaccination" to align with the most current guidelines. While acknowledging the importance of the measure concept, MAP raised concerns that the measure was not suitable for use in the MIPS program due to the ability of providers to choose which measures to report. MAP members noted that the measure would not provide meaningful surveillance or performance data as only providers with high vaccination rates in their patient population would choose to report the measure."
Is the measure currently used in a CMS program	N/A

MERIT Submission Information MUC2023 164	Description
If previously used, please describe the history of the measure in CMS program	N/A
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	N/A
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
History of endorsement	N/A
Path to endorsement	Unknown
Measure Specification Details	
Measure Description	Percentage of patients aged 18 years and older seen for a visit during the performance period that are up to date on their COVID-19 vaccinations as defined by CDC guidelines on current vaccination.
Data source	Registries
Level of analysis	Clinician: Individual and Group (MIPS-Quality only)
Numerator	Patients that are up to date on their COVID-19 vaccinations as defined by CDC guidelines on current vaccination as of the date of the encounter.
Denominator	All patients aged 18 years and older seen for a visit during the performance period.
Numerator exclusions	N/A
Denominator exclusions	Patient received hospice services any time during the performance period.
Denominator exceptions	Patients that are not up to date on their COVID-19 vaccinations as defined by CDC guidelines on current vaccination because medical contraindication documented by clinician.
Risk adjustment	No
Development Status	Not Fully Developed
If not fully developed, development stage	Specification
Target population	All-payer
Measure type	Process
Is the measure a composite or component of a composite?	No

MERIT Submission Information MUC2023 164	Description
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.13.2. MUC2023-164 Measure Evaluation

MUC2023 164 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	Widespread vaccination to prevent severe COVID-19 infection will be critically important to stemming the morbidity and mortality caused by this disease. In March 2023, the Kaiser Family Foundation reported that merely a quarter of adults (23%) reported having received the bivalent booster.[69]	--	The study population is the same as the target quality program population.

[69] Kaiser Family Foundation. COVID 19 Vaccine Monitor Dashboard. https://ww w.kff.org/coronavirus-covid-19/dashboard/kff-covid-19-vaccine-monitor-dashboard/

MUC2023-164 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	During face validity testing, the denominator, exclusions, and exceptions were largely supported by interview participants and align with intent of use.	--	Most persons and entities in the quality program population are included in the specification.
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	Some data elements are in defined fields in electronic sources.	--	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden to data collection and reporting? (*Context of Use)*	Performance improvement has achievable and relevant benchmark of care balanced with burden of data collection.	--	All the performance improvements to the benchmark have a significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound? *(Context of Use)*	Reliability signal-to-noise (median Statistical results) was 0.986. The 10th percentile reliability score was 0.849, the 25th percentile was 0.947, the median was 0.986, the 75th percentile was 0.994, and the 90th percentile was 0.997. Reliability coefficients above 0.70 are considered sufficient to draw conclusions about groups, and values above 0.9 are considered	--	Most or all entities have reliability above the threshold (0.60) within the quality program population.

MUC2023-164 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	sufficient to draw conclusions about individuals.		
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	USPSTF grade provided was Grade A (i.e., strong recommendation or similar). ACIP guidelines support the measure, all of which are evidence-based. On February 4, 2022, ACIP issued a standard recommendation for use of the Moderna COVID-19 vaccine in persons aged >18 years. Face validity was assessed in two phases. Phase 1 interviewed six clinicians. Three out of six interviewees indicated agreement that better performance was indicative of better care. Phase 2 interviewed four clinicians and two registry staff.	The three interview participants noted concerns that 1) some patient populations would be less likely to get vaccinated experts and patients/caregivers, and 2) data may not be reliable from some practices. Further, they noted concerns that changes in guidelines would affect implementation of the measure.	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population.
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	--	Risk adjustment or stratification not reported.	N/A

MUC2023-164 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	There is an explicit articulation of the resources and context that might facilitate or be a barrier to the way an entity may improve.	--	There is an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-164 Simulated Measure Reliability Tables

The performance score is a percentage of patients that are up to date on COVID-19 vaccination for each entity.

Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across the 33 entities. σ^2_{within} is the variance (standard deviation squared) of the score within a single entity. The total number of patients across all entities is 26,424 with each entity having at least 11 patients.

The measure report indicates a median signal-to-noise reliability of 0.986.

Simulated decile tables:

Computer simulation was used to create a dataset that mirrors, as closely as possible, the mean, standard deviation, and percentile information provided for the performance score and calculated reliability. Tables 3.13.3 and 3.13.4 are created from the simulated dataset and provide reviewers with a more standardized format to assess reliability.

For Table 3.13.3, entities were sorted by performance score, and the average score by decile (estimated from the simulated data) is reported along with the number of entities included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.13.3. Importance (Decile by performance score)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	48.03 (20.74)	13.40	17.40	26.04	36.25	43.84	48.23	54.13	56.14	61.23	69.24	89.22	92.91
Entities	33	1	4	4	4	3	3	3	3	3	3	3	1

For Table 3.13.4, entities were sorted by reliability and the average reliability by decile (estimated from the simulated data) is reported along with the number of entities included in each average. Average, minimum, and maximum reliability and expected events are also included.

Table 3.13.4. Reliability (Decile by reliability)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Reliability	0.95	0.70	0.80	0.90	0.96	0.98	0.98	0.98	0.99	1.00	1.00	1.00	1.00
Entities	33	1	4	4	4	3	3	3	3	3	3	3	1

Assumptions:

The information provided for the measure score and reliability was based on 33 entities and 26,424 total patients.

Interpretation:

The reported median reliability is 0.986. The reliability simulation suggests that few, if any, of the entities have reliability below 0.6.

3.14 MUC2023-211 Melanoma: Tracking and Evaluation of Recurrence

Description: Percentage of patients who had an excisional surgery for melanoma or melanoma in situ with initial AJCC staging of 0, I, or II, in the past 5 years in which the operating provider examines and/or diagnoses the patient for recurrence of melanoma.

Measure Type: Outcome

Level of Analysis: Clinician: Individual and Group

Data Source(s): Electronic Health Record; Registries

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Melanoma: Tracking and Evaluation of Recurrence measure to the MIPS quality measure set. This measure has been in use within MIPS as a QCDR measure since 2022: AAD 14: Melanoma: tracking and evaluation of recurrence. This measure will evaluate the frequency of recurrence along with the type of recurrence after an excisional procedure and aims to drive communication about the recurrence status of melanoma patients, as a lack of communication has been recognized between the excising provider and provider continuing care. This measure will allow for the development of a system in which melanomas can be accurately tracked to increase the understanding of the effectiveness of care. CMS is considering adding this quality measure to the MIPS program to fill a gap within the dermatologic clinical topic and there is potential to replace the current MIPS 137: Melanoma: Continuity of Care with this more robust outcome measure. While the Dermatological MIPS specialty measure set includes 13 measures, only 6 are specialty specific.

Table 3.14.1. MUC2023-211 Brief Summary of Measure Information

MERIT Submission Information MUC2023-211	Description
Measure name	Melanoma: Tracking and Evaluation of Recurrence
MUC ID	MUC2023-211
Cascade priority	Wellness and Prevention
Measure steward	American Academy of Dermatology
Measure Developer	American Academy of Dermatology

MERIT Submission Information MUC2023-211	Description
Program submitted to	Merit-based Incentive Payment System-Quality
Committee assigned to	Clinician Committee
Related measures in the program	N/A
Is this a new measure in this year's MUC List?	Measure currently used in a CMS program being submitted as-is for a new or different program
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program	Yes
If previously used, please describe the history of the measure in CMS program	This measure is currently used in the Merit-based Incentive Payment System (MIPS). It has been used since 2022 and is being submitted as is, aligning with the current version.
Any other program the measure is in use	Merit-based Incentive Payment System
Is this measure being proposed to meet a statutory requirement?	N/A
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
History of endorsement	N/A
Path to endorsement	N/A
Measure Specification Details	
Measure Description	Percentage of patients who had an excisional surgery for melanoma or melanoma in situ with initial AJCC staging of 0, I, or II, in the past 5 years in which the operating provider examines and/or diagnoses the patient for recurrence of melanoma.
Data source	Electronic Health Record; Registries
Level of analysis	Clinician: Individual and Group (MIPS-Quality only)
Numerator	Numerator Criteria 1: Documentation by the provider who performed the surgery that an exam for recurrence of melanoma was performed on the patient within the reporting period. Numerator Criteria 2: (Reported Score) All patients that were diagnosed with a recurrent melanoma in the current reporting year.

MERIT Submission Information MUC2023-211	Description
Denominator	Denominator Criteria 1 & 2: All patients that the provider has performed a type of excisional surgery for melanoma or melanoma in situ in the past 5 years with an initial AJCC staging of 0, I, or II.
Numerator exceptions	N/A
Denominator exclusions	Documentation that the patient is deceased. Documentation of patient refusal of examination. Documentation that the patient was lost to follow-up (documentation must include information that the provider was unable to reach the patient by phone, mail or secure electronic mail – at least one method must be documented).
Denominator exceptions	N/A
Risk adjustment	No
Development Status	Fully Developed
If not fully developed, development stage	N/A
Target population	All payer
Measure type	Outcome
Is the measure a composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.14.2. MUC2023-211 Measure Evaluation

MUC2023-211 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	Evidence suggests that a combination of patient-led surveillance and routine follow-up visits may improve melanoma detection.[70] Additional studies discuss the psychosocial factors related to follow-up examinations as well as the impact on patient well-being, potential adherence to schedules, and clinician practice procedure.[71] The authors also note that explicit guidance and specialist support for clinicians would benefit patient outcomes. There are limited guidelines regarding an established follow-up cadence, as well as appropriate tracking procedures related to melanoma care and recurrence.	--	The study population is the same as the target quality program population.
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	Denominator criteria is clearly defined: All patients for which the provider has performed a type of excisional surgery for melanoma or melanoma in situ in the past 5 years with an initial AJCC staging of 0, I, or II. Exclusions include refusal of examination, patient being deceased, and/or patient was lost to follow-up.	--	Most persons and entities in the quality program population are included in the specification.

[70] Dretzke, J., Chaudri, T., Balaji, R., Mehanna, H., Nankivell, P., Moore, D. J., & PETNECK2 Research Team (2023). A systematic review of the effectiveness of patient-initiated follow-up after cancer. Cancer medicine, 12(18), 19057–19071. https://doi.org/10.1002/cam4.6462

[71] Rychetnik, L., McCaffery, K., Morton, R., & Irwig, L. (2013). Psychosocial aspects of post-treatment follow-up for stage I/II melanoma: a systematic review of the literature. Psycho-oncology, 22(4), 721–736. https://doi.org/10.1002/pon.3060

MUC2023-211 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	Some data elements are captured within defined fields via electronic sources. Feasibility testing assessed availability and retrievability of numerator data as well as clinician burden associated with collecting data. Members of the AAD Steering Committee (n=11) were invited to participate. Data elements ranged from anywhere between 60-100% of the sites included in the assessment. On a Likert Scale from 1 to 5 designed to assess burden (higher numbers indicate higher perceived burden), numerator data elements ranged from 0 (n/a) to 3.5. The highest burden was associated with the 'Patient deceased' data element.	--	Some entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden to data collection and reporting? (*Context of Use)*	A structured follow-up assessment cadence could improve likelihood of earlier detection of recurrent disease, allowing for advantageous treatment.	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound? *(Context of Use)*	A Random Split-Half Correlation analysis was conducted (n=7) and resulted in a 0.21 correlation. This is below the developer's stated desired 0.4 threshold. However, several factors impacting reliability were noted (e.g., small sample affected the between dermatologist variation).	It is unlikely that the reliability for more than a few, if any, entities is greater than 0.6.	No entities have reliability above the threshold (0.60) within population that extrapolates to the quality program population.
Validity: May providers/facilities/care systems effectively improve on this measure?	Face validity voting indicated 15 out of 16 members agreed with face validity of the measure.	Outcome of additional follow-up for patients with melanoma that is perceived as intensive surveillance is not determined.	There is an association between the entity and the measure focus in a population that

MUC2023-211 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
(Context of Use)	USPSTF Grade B, Moderate recommendation or similar. The measure is based on the National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology-Melanoma (2016).		extrapolates to the quality program population.
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	The measure is not risk adjusted.	The developer indicated that social risk factors were analyzed during testing but were determined not to be reliably documented or abstracted and was not included in the testing report.' Further, 'data were not available to evaluate risk adjustment or stratification.'	N/A
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	--	There is not an explicit articulation of the resources and context that might facilitate or be a barrier to the way an entity may improve.	Unable to determine if there is an articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-211 Measure Reliability

The performance score is the proportion of patients who had surgery for melanoma in the past 5 years in which the operating provider examined and/or diagnosed the patient for recurrence of melanoma. For the reliability testing method & reliability score information provided, simulated reliability tables were not developed.

The measure report indicates a random split half correlation of 0.21.

Interpretation:

The reported random split-half reliability is 0.21. It is unlikely that the reliability for more than a few, if any entities is greater than 0.6.

Part C & D Star Rating Medicare Program

3.15 MUC2023-137 Initial Opioid Prescribing for Long Duration (IOP-LD)

Description: The IOP-LD measure analyzes the percentage of Medicare Part D beneficiaries, 18 years or older, with at least one initial opioid prescription for more than 7 cumulative days' supply.

Measure Type: Process

Level of Analysis: Health Plan

Data Source(s): Administrative Data (non-claims); Claims Data; Prescription Drug Event (PDE)

Development Status: Fully Developed

Endorsement Status: Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering moving the Initial Opioid Prescribing for Long Duration (IOP-LD) measure from the Part C & D Display page to the Part C & D Star Ratings program measure set.

The measure will fill a gap area in opioid safety for patients who are opioid naïve but have been prescribed opioid prescription claims for more than 7 cumulative days' supply. The duration of initial opioid exposure is associated with a higher likelihood for long-term opioid use, which is linked to greater risks of abuse and overdose. The literature is consistent in finding that greater days' supply for initial opioid prescriptions is associated with significant harms, including increased risk of long-term opioid use, opioid misuse, and overdose. The IOP-LD measure is intended for use for retrospective population-level performance measurement and is not intended to guide clinical decision-making for individual patients.

Most recent data on average for the IOP-LD rate among MA-PDs was about 15% while for PDPs it was about 16%.

The IOP-LD measure has been publicly available on the Part C & D display page since Fall 2022 (data from measurement year 2021).

Table 3.15.1. MUC2023-137 Brief Summary of Measure Information

MERIT Submission Information MUC2023-137	Description
Measure name	Initial Opioid Prescribing for Long Duration (IOP-LD)
MUC ID	MUC2023-137
Cascade priority	Safety
Measure steward	Pharmacy Quality Alliance
Measure Developer	Centers for Medicare & Medicaid Services
Program submitted to	Part C & D Star Rating [Medicare]
Committee assigned to	Clinician Committee
Related measures in the program	N/A
Is this a new measure in this year's MUC List?	Yes
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program	Part C & D Display Page (Medicare)
If previously used, please describe the history of the measure in CMS program	Measure currently used in a CMS program being submitted as-is for a new or different program
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	N/A
CBE endorsement status	Endorsed
CBE endorsement number if applicable	CBE 3558
History of endorsement	Most recent CDP endorsement: 2019
Path to endorsement	Unknown
Measure Specification Details	
Measure Description	The IOP-LD measure analyzes the percentage of Medicare Part D beneficiaries, 18 years or older, with at least one initial opioid prescription for more than 7 cumulative days' supply.
Data source	Administrative Data (non-claims); Claims Data; Prescription Drug Event (PDE)
Level of analysis	Health Plan

MERIT Submission Information MUC2023-137	Description
Numerator	The number of member-years of beneficiaries in the denominator with more than 7 cumulative days' supply for opioid prescription claims within any 3-day opioid initiation period.
Denominator	The number of member-years of enrolled beneficiaries, 18 years or older with 1 or more opioid prescription claim(s) with a negative medication history during the 90-days lookback period.
Numerator exclusions	N/A
Denominator exclusions	Beneficiaries with a cancer diagnosis, sickle cell diagnosis, enrolled in hospice, or in palliative care at any time during the measurement period or the 90 days prior to the index prescription start date are excluded from the denominator.
Denominator exceptions	N/A
Risk adjustment	N/A
Development Status	Fully Developed
If not fully developed, development stage	N/A
Target population	Medicare Part D: Medicare Advantage Prescription Drug Plans and Prescription Drug Plans
Measure type	Process
Is the measure composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.15.2. MUC2023-137 Measure Evaluation

MUC2023-137 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	This measure supports HHS and CMS priorities to prevent opioid-related overdoses and deaths and to promote safer use of opioids among the Medicare Part D population. Additionally, adequate assessment and management of pain is important after opioid initiation to minimize risk of long-term opioid use, opioid misuse, and overdose. This measure was endorsed by NQF in 2019 (NQF #3558) and is aligned with the CMS Behavioral Health Strategy and CMS Meaningful Measures 2.0 Safety goals. The literature is fully consistent in finding that greater days' supply for initial prescriptions is associated with significant harms, including increased risk of long-term opioid use, opioid misuse, and overdose. The higher duration of initial opioid exposure is associated with a higher likelihood for long-term opioid use, which is linked to greater risks of abuse and overdose. The 2016 Centers for Disease Control and Prevention (CDC) Guideline for Prescribing Opioids for Chronic Pain recommends that when opioids are used for acute pain (i.e., pain with abrupt onset and caused by an injury or other process that is not ongoing), no greater quantity should be prescribed than is needed for the expected duration of pain severe enough to require opioids. According to the 2016 guideline, 3 days or less will often be sufficient and more than 7 days will rarely be needed.[72] Numerous studies present findings that longer duration of	Evidence from the literature is not reviewed beyond the CDC recommendation.	The study population is the same as the target quality program population

[72] Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain - United States, 2016. *MMWR Recomm Rep*. Mar 18 2016;65(1):1-49. doi:10.15585/mmwr.rr6501e1

MUC2023-137 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	use of opioids is associated with a higher incidence for prolonged opioid use, and subsequent misuse.[73, 74, 75, 76]		
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	Denominator exclusions are consistent with 2022 CDC Guideline for prescribing opioids for chronic pain, i.e., persons near end of life (hospice, palliative care) or who have severely painful chronic illnesses (cancer, sickle cell) are excluded.	--	Most persons and entities in the quality program population are included in the specification.
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	All data elements are available in electronic claims records. There is no provider burden associated with the measure.	A potential unintended consequence is that individuals entering Long-Term Care after receiving opioids while in an inpatient setting (which would not be captured via Part D prescription claims) may appear to have a negative medication history and inadvertently be included in the IOP-LD denominator.	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Do the benefits of performance improvement to the	Overall mean performance score: 18.5%; minimum, 0%; 10th, 8.14%; median, 15.04%; 90th, 27.58; maximum, 91.75. Scores demonstrate wide variation in performance and ample opportunity for improvement.	--	Most of the performance improvements to the benchmark have a

[73] Tehrani AB, Henke RM, Ali MM, Mutter R, Mark TL. Trends in average days' supply of opioid medications in Medicaid and commercial insurance. *Addict Behav*. Jan 2018;76:218-222. doi:10.1016/j.addbeh.2017.08.005

[74] Shah A, Hayes CJ, Martin BC. Characteristics of Initial Prescription Episodes and Likelihood of Long-Term Opioid Use - United States, 2006-2015. *MMWR Morb Mortal Wkly Rep*. Mar 17 2017;66(10):265-269. doi:10.15585/mmwr.mm6610a1

[75] Shah A, Hayes CJ, Martin BC. Factors Influencing Long-Term Opioid Use Among Opioid Naive Patients: An Examination of Initial Prescription Characteristics and Pain Etiologies. *J Pain*. Nov 2017;18(11):1374-1383. doi:10.1016/j.jpain.2017.06.010

[76] Zhang Y, Johnson P, Jeng PJ, et al. First Opioid Prescription and Subsequent High-Risk Opioid Use: a National Study of Privately Insured and Medicare Advantage Adults. *J Gen Intern Med*. Dec 2018;33(12):2156-2162. doi:10.1007/s11606-018-4628-y

MUC2023-137 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
achievable benchmark of care exceed the burden to data collection and reporting? (*Context of Use)*	There is evidence of a performance gap among the MA-PDs and PDPs evaluated using the IOP-LD measure. Based on CY2021 CMS Part D Patient Safety Report data using contracts with N>=30, average IOP-LD rates (SD) [IQR] for MA-PDs 18.9% (14.4%) [7.0%] and PDPs 14% (4.2%) [5.0%] demonstrate substantial variance in performance and room for improvement in the IOP-LD measure and continual monitoring of beneficiaries newly prescribed of opioid prescriptions.		significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound? *(Context of Use)*	The reliability for the IOP-LD measure was 0.95 for MAPD and 0.94 for PDPs.	--	Most or all entities have reliability above the threshold (0.60) *within* the quality program population
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	In a face validity evaluation, 18 of 20 voters "recommend the IOP-LD measure be considered for endorsement by the QMEP." From the CDC Clinical Practice Guideline for Prescribing Opioids for Pain – United States, 2022 - "When opioids are needed for acute pain, clinicians should prescribe no greater quantity than needed for the expected duration of pain severe enough to require opioids [recommendation category: A; evidence type: 4].[77] Data suggest that pain improves within days for many patients with common types of acute pain in primary care or emergency department settings. Analysis of nationwide U.S. commercial insurance claims in 2014 found median durations of initial opioid analgesic prescriptions for acute pain indications in primary care settings were 4–7 days, suggesting that in most cases, clinicians considered an initial opioid prescription of 4–7	No details were provided regarding the specific criteria used for voting on face validity (e.g., low, moderate, or high confidence that the measure could distinguish high vs. low performing plans).	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population.

[77] Dowell D, Ragan KR, Jones CM, Baldwin GT, Chou R. CDC Clinical Practice Guideline for Prescribing Opioids for Pain — United States, 2022. MMWR Recomm Rep 2022;71(No. RR-3):1–95. DOI: http://dx.doi.org/10.15585/mmwr.rr7103a1

MUC2023-137 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	days' duration sufficient.[78] [USPSTF Grade A, strong recommendation] In addition to face validity assessment by the PQA Quality Metrics Expert Panel, the measure was reviewed by PQA's Measure Validity Panel (MVP) as part of PQA's standard, transparent measure development process. The MVP is made up of an independent group of individuals not involved in the development or review of the measure concept or draft measure. Through discussion and vote, the MVP determines whether the performance measure scores have face validity. Of the 7 MVP members who voted, 100% agreed or strongly agreed that the scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality between health plans.		
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	The measure is not risk adjusted.	--	N/A
Usability: Is there opportunity for improvement on this measure in the intended use setting? **(Context of Use)**	71 out of 82 measured entities who responded to the question thought the information was easy to understand and useful for decision-making. This measure is currently in use for CMS Parts C & D. CMS Part D does not expect a zero-percentage measure rate for	--	There is an explicit articulation of the resources and context that might facilitate improvement within

[78] Mundkur ML, Franklin JM, Abdia Y, et al. Days' supply of initial opioid analgesic prescriptions and additional fills for acute pain conditions treated in the primary care setting—United States, 2014. MMWR Morb Mortal Wkly Rep 2019;68:140–3. https://doi.org/10.15585/mmwr.mm6806a3 PMID:30763301

MUC2023-137 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	the IOP-LD. CMS Part D opioid guidance provides flexibility for longer durations. The criteria usability of the IOP-LD measure is a plan-level measure (is not intended to be a pharmacy, physician, or hospital level measure). The IOP-LD measure is intended for use for retrospective population-level performance measurement and is not intended to guide clinical decision-making. The IOP-LD measure is an additional tool for Part D plans to monitor initial opioid prescriptions to prevent an increased risk for chronic opioid use and opioid use disorder. Access to medications for opioid use disorder are not impeded by the measure specifications.		the quality program population

MUC2023-137 Measure Reliability

The performance score is the percentage of events per total opportunities for each entity.

The reported mean reliability is 0.95 for MAPD and 0.94 for PDPs as provided by the developer.

Interpretation:

The reported mean reliability is 0.95 for MAPD and 0.94 for PDPs. An additional assessment of the reliability by denominator decile would demonstrate whether the reliability falls below 0.6 for entities with a denominator close to the minimum (30 member-years).

3.16 MUC2023-179 Initiation and Engagement of Substance Use Disorder Treatment (IET)

Description: The percentage of new substance use disorder (SUD) episodes that result in treatment initiation and engagement.

Measure Type: Process

Level of Analysis: Administrative Data (non-claims); Claims Data; Electronic Clinical Data (non-EHR); Electronic Health Record

Data Source(s): Health Plan

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the Initiation and Engagement of Substance Use Disorder Treatment (IET) measure to the Part C & D Star Ratings program. This measure was formerly called the Initiation and Engagement of Alcohol and other Drug Dependence Treatment measure. This measure supports behavioral health, a high priority area for the program. The addition of this measure would support CMS's efforts to implement the Universal Foundation set of measures across CMS quality programs as this measure is part of the core set of measures.

This measure has the potential to fill a gap by focusing on whether Medicare Advantage (MA) enrollees are getting evidence-based treatment for new substance use disorder (SUD) episodes.

Despite known and effective treatments, in the most recent data on average only 36% of MA enrollees with new SUD episodes are initiating treatment through an inpatient SUD admission, outpatient visit, intensive outpatient encounter, partial hospitalization, telehealth visit or medication treatment within 14 days.

Early treatment engagement is a critical step between accessing care and completing a full course of treatment. Individuals who engage in early SUD treatment have been found to have decreased odds of negative outcomes, including mortality. In the most recent data on average 5% of MA enrollees with new SUD episodes have evidence of treatment engagement within 34 days of initiation.

This measure has been on the CMS display page at cms.gov since 2014 Star Ratings (2012 measurement year).

Table 3.16.1. MUC2023-179 Brief Summary of Measure Information

MERIT Submission Information MUC2023-179	Description
Measure name	Initiation and Engagement of Substance Use Disorder Treatment (IET)
MUC ID	MUC2023-179
Cascade priority	Behavioral Health
Measure steward	National Committee for Quality Assurance (NCQA)
Measure Developer	Centers for Medicare & Medicaid Services
Program submitted to	Medicare Part C & D Star Rating
Committee assigned to	Clinician Committee
Related measures in the program	N/A
Is this a new measure in this year's MUC List?	Yes
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program	N/A
If previously used, please describe the history of the measure in CMS program	New measure never reviewed by Measure Applications Partnership (MAP) Workgroup or used in a CMS program
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	N/A
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
Measure Specification Details	
Measure Description	The percentage of new substance use disorder (SUD) episodes that result in treatment initiation and engagement. Two rates are reported.

MERIT Submission Information MUC2023-179	Description
	Initiation of SUD Treatment: The percentage of new SUD episodes that result in treatment initiation through an inpatient SUD admission, outpatient visit, intensive outpatient encounter, partial hospitalization, telehealth visit or medication treatment within 14 days. (Presented in this MERIT submission form.) Engagement of SUD Treatment: The percentage of new SUD episodes that have evidence of treatment engagement within 34 days of initiation. The definition of an episode follows the NCQA HEDIS specification for this measure.
Data source	Administrative Data (non-claims); Claims Data; Electronic Clinical Data (non-EHR); Electronic Health Record
Level of analysis	Health Plan
Numerator	SUD Treatment initiation within 14 days of a new SUD episodes date (for alcohol use disorder, opioid use disorder, or other substance use disorder) in an outpatient, intensive outpatient, partial hospitalization, observation, telehealth or emergency department for members ages 13 and older.
Numerator Exclusions	Exclude initiation events that occur on the same day as the SUD Episode Date if with the same provider except for medication treatment dispensing events and medication administration events.
Denominator	New SUD episodes (for alcohol use disorder, opioid use disorder, or other substance use disorder) during the Intake Period in an outpatient, intensive outpatient, partial hospitalization, observation, telehealth or emergency department for members ages 13 and older.
Numerator exceptions	N/A
Denominator exclusions	Exclude the member from the denominator if the initiation of treatment event is an inpatient stay with a discharge date after November 27 of the measurement year. Exclude members in hospice or using hospice services anytime during the measurement year.
Denominator exceptions	N/A
Risk adjustment	No
Development Status	Fully Developed
If not fully developed, development stage	N/A
Target population	Medicare Advantage enrollees

MERIT Submission Information MUC2023-179	Description
Measure type	Process
Is the measure a composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.16.2. MUC2023-179 Measure Evaluation

MUC2023-179 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Importance: Does the measure align with goals and priorities?	Evidence-based treatment for SUD includes both psychosocial supports and, for opioid and alcohol use disorders, medication.[79, 80, 81, 82]	--	While the study population differs from the target quality program population, the importance for the

[79] Kampman, K., K. Freedman. 2020. "American Society of Addiction Medicine (ASAM) National Practice Guideline for the Treatment of Opioid Use Disorder: 2020 Focused Update." Journal of Addiction Medicine 14, no. 2S: 1–91, https://doi.org/10.1097/ADM.0000000000000633.
[80] Reus, V. et al. 2018. "Practice Guideline for the Pharmacological Treatment of Patients with Alcohol Use Disorder." American Journal of Psychiatry 175(1), 86–90. doi:10.1176/appi.ajp.2017.1750101
[81] Department of Veteran Affairs, Department of Defense. 2015. VA/DoD Clinical Practice Guideline for the Management of Substance Use Disorders. Washington DC: Department of Veterans Affairs, Department of Defense.
[82] Michigan Quality Improvement Consortium. August 2015. Screening, diagnosis and referral for substance use disorders. Southfield (MI): Michigan Quality Improvement Consortium.

MUC2023-179 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
(*Concept of Interest)*	Less than 20% of individuals with SUD receive specialty care.[83, 84] Individuals who engage in early SUD treatment have been found to have decreased odds of negative outcomes, including mortality.[85, 86]		selected program population can be extrapolated.
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	Measure reports two rates: IET Initiation and IET Engagement. Denominator exclusions appear appropriate (hospice during measurement year, SUD episodes that begin too late to be fully within the measurement year, i.e., beginning after Nov. 27). For measurement year 2022, the specifications were updated: measure changed from member-based to episode-based; lengthened the negative SUD history period from 60 to 194 days to limit the number of	No patient/encounter level reliability or validity testing was performed.	Most persons and entities in the quality program population are included in the specification.

[83] Substance Abuse and Mental Health Services Administration (SAMHSA). 2019. Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health. https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.htm#mhisud

[84] Williams, E.C., T.E. Matson, & A.H. Harris. 2019. "Strategies to Increase Implementation of Pharmacotherapy for Alcohol Use Disorders: A Structured Review of Care Delivery and Implementation Interventions." Addiction Science & Clinical Practice 14(1), 6.
[85] Paddock, S.M., K.A. Hepner, T. Hudson, et al. 2017. "Association Between Process-Based Quality Indicators and Mortality for Patients with Substance Use Disorders." J Stud Alcohol Drugs 78:588–96.
[86] Watkins, K.E., S.M. Paddock, T.J. Hudson, et al. 2016. "Association Between Quality Measures and Mortality in individuals with Co-Occurring Mental Health and Substance Use Disorders." J Subst Abuse Treat 69:1–8.

MUC2023-179 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	members receiving ongoing treatment who inadvertently fall into the denominator; removed ED visits and medically managed withdrawal services from the negative SUD history period (since these events are not suggestive of ongoing or planned treatment); and removed the requirement that a psychosocial treatment encounter accompany pharmacotherapy (to align with most current clinical practice guidelines).		
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	No unintended consequences were identified. All data elements are in defined fields in electronic sources.	IET Initiation and Engagement: One expert opposed face validity based on complexities and the burden of the revised episode structure in denominator.	Most entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden to data collection and reporting? (*Context of Use)*	In 2018, 20.3 million individuals in the U.S. 12 years of age or older (approximately 7.4% of the population) were classified as having an SUD within the past year.[87] Individuals with SUDs are at increased risk of overdose, injury, soft tissue infections and mortality.[88] In 2017, drug	Patients/caregivers did not provide input on the measure.	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.

[87] Substance Abuse and Mental Health Services Administration (SAMHSA). 2019. Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health. https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.htm#mhisud

[88] Bahorik, A.L., D.D. Satre, A.H. Kline-Simon, C.M. Weisner, C.L. Campbell. (2017). Alcohol, Cannabis, and Opioid Use Disorders, and Disease Burden in an Integrated Health Care System. J Addiction Medicine 11(1),3–9.

MUC2023-179 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	overdose accounted for more than 70,200 deaths—67.8% due to opioid use.[89] In 2016, 1 in 10 deaths among working adults in the U.S. were attributed to alcohol misuse.[90] The measure potentially impacts 868,407 Medicare Advantage enrollees annually. IET Initiation: Wide range in performance scores and opportunity for improvement; n=529, mean 33.01%, min 3.65, 10th 15.94, median 33.09, 90th 47.53, max 85.42. Evidence for gap by social risk factors: Overall, beneficiaries with dual eligibility or low-income subsidy (DE/LIS) were more likely than others to have initiated treatment within 14 days of a SUD diagnosis.[91] IET Engagement: Wide range in performance scores and opportunity for improvement; n=529, mean 5.05%, min 0.0, 10th 1.25, median 4.50, 90th 9.80, max 21.43. Evidence for gap by social risk factors: Overall, DE/LIS beneficiaries who		

[89] NIDA. 2018b. Overdose Death Rates. https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates
[90] U.S. Department of Health and Human Services (HHS), Office of the Surgeon General. November 2016. Facing Addiction in America: The Surgeon General's Report on Alcohol, Drugs, and Health. Washington, DC: HHS.
[91] Martino, SC, Elliott, MN, Dembosky, JW, Hambarsoomian, K, Klein, DJ, Gildner, J, and Haviland, AM. Disparities in Health Care in Medicare Advantage Associated with Dual Eligibility or Eligibility for a Low Income Subsidy. Baltimore, MD: CMS Office of Minority Health. 2021.

MUC2023-179 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	initiated treatment were more likely than others to have had two or more additional services within 30 days of their initial visit for treatment.[92] API, Black, and Hispanic beneficiaries were less likely than Whites to have this, women were less likely than men to have this, and rural residents were more likely than urban to have this.		
Reliability: Is measure performance scientifically sound? *(Context of Use)*	IET Initiation: Signal-to-noise analysis was performed (n=408): minimum, 0.647; 20th percentile, 0.892; median, 0.968; 80th percentile, 0.991; maximum, 1.0; more than 80% of health plans have reliability above 0.60. IET Engagement: Signal-to-noise analysis was performed (n=408): minimum, 0.174; 20th percentile. 0.714; median, 0.901; 80th percentile, 0.979; maximum, 1.0; more than 80% of health plans have reliability above 0.60.	IET Initiation and Engagement: Fewer than 20% of health plans have reliability below 0.60. Measure reliability and validity were tested prior to the measure updates for measurement year 2023, specifically, before the change to episode level; however, the submission indicates that expert feedback and unspecified additional testing show that changes did not significantly impact intent or performance.	Most or all entities have reliability above the threshold (0.60) *within* the quality program population.
Validity: May providers/facilities/care systems effectively improve on this measure?	IET Initiation: Construct validity was tested using Pearson correlation, comparing IET initiation with Follow-up for ED visit for SUD 7-day (0.24; weak) and 30-day (0.26, weak) measures, and with the IET	IET Initiation: weak correlations with related measures (FUA 7-day and FUA 30-day). See above regarding measure validity and reliability being tested	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population.

[92] Martino, SC, Elliott, MN, Haas, A, Klein, DJ, Hambarsoomian, K, Haviland, AM, Adams, JL, Weech-Maldonado, R, Gildner, JL, Edwards, CA, and Dembosky, JW. Trends in Racial, Ethnic, Sex, and Rural-Urban Inequities in Health Care in Medicare Advantage: 2009–2018. Baltimore, MD: CMS Office of Minority Health. 2021.

MUC2023-179 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
(Context of Use)	engagement indicator (0.59, p<.0001; moderate) (n=408). IET Engagement: Construct validity was tested using Pearson correlation, comparing IET initiation with Follow-up for ED visit for SUD 7-day (0.39, p<.0001) 30-day (0.41, p<.0001) measures, and with the IET initiation indicator, (0.59, p<.0001; moderate) (n=408). IET Initiation and Engagement: In face validity testing, 10 of 11 experts agreed that the measure could differentiate between facilities on care quality. Guideline statements that most closely align with measure concept: (1) Because many substance use disorders are chronic, patients usually require long-term treatment, although the intensity and specific components of treatment may vary over time [I]; (2) The duration of treatment should be tailored to the individual patient's needs and may vary from a few months to several years [I]; (3) It is important to intensify the monitoring for substance use during periods when the patient is at a high risk of relapsing, including during the early stages of treatment, times of transition to less intensive levels of care, and the first year after active treatment has	before specifications were updated.	

MUC2023-179 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
	ceased [I].[93] The USPSTF grade for guideline (3) is Grade A, Strong recommendation Similar guidelines from APA, American Society of Addiction Medicine, Michigan Quality Improvement Consortium, Department of Veteran Affairs, Department of Defense, and USPSTF were cited.[94, 95, 96, 97, 98, 99]		
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	Not risk adjusted Stratification is recommended for race, ethnicity, age group (13-17, 18-64, 65+), diagnosis cohort (alcohol, opioid, other substance); stratification approach is based on the disparities identified in the literature review.	--	N/A

[93] Kleber HD, et al. (2006). Treatment of patients with substance use disorders, second edition. American Journal of Psychiatry, 164(4),1–276.
[94] Reus, V. et al. (2018). Practice Guideline for the Pharmacological Treatment of Patients with Alcohol Use Disorder. American Journal of Psychiatry, 175(1), 86-90. doi:10.1176/appi.ajp.2017.1750101
[95] Kampman, K., Freedman, K. (2020). American Society of Addiction Medicine (ASAM) National Practice Guideline for the Treatment of Opioid Use Disorder: 2020 Focused Update. Journal of Addiction Medicine; 14, no. 2S: 1–91, https://doi.org/10.1097/ADM.0000000000000633.
[96] Kampman, K., Jarvis, M. (2015). American Society of Addiction Medicine (ASAM) National Practice Guideline for the Use of Medications in the Treatment of Addiction Involving Opioid Use. Journal of Addiction Medicine; 9(5): 358–367. DOI: 10.1097/ADM.0000000000000166.
[97] Michigan Quality Improvement Consortium. Screening, diagnosis and referral for substance use disorders. Michigan Quality Improvement Consortium; 2021 July. 1 p.
[98] Department of Veteran Affairs, Department of Defense. (2015). VA/DoD Clinical Practice Guideline for the Management of Substance Use Disorders. Washington DC: Department of Veterans Affairs, Department of Defense.
[99] Final Recommendation Statement: Alcohol Misuse: Screening and Behavioral Counseling Interventions in Primary Care. U.S. Preventive Services Task Force. May 2013. https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/alcohol-misuse-screening-and-behavioral-counseling-interventions-in-primary-care

MUC2023-179 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	Plans may use these data to target education and outreach efforts and strengthen patient access to care. Measure is currently in use in four other CMS programs: Marketplace Quality Rating System (MQRS); Medicaid: Adult Core Set (MACS); Medicaid: Home Health Core Set (MHHCS); Merit-based Incentive Payment System (MIPS).	Submission is not specific about how health plan will reach clinicians. Measured entities did not provide input on the measure.	There is an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-179 Simulated Measure Reliability Tables

The performance score is a percentage of SUD episodes that result in treatment initiation and engagement.

Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across the entities. σ^2_{within} is the variance (standard deviation squared) of the score within a single entity. The study includes 408 entities but does not report the total number of patients included. The estimated annual denominator size, i.e., total number of patients, is 868,407 across an estimated 3,998 entities.

The measure report indicates a median signal-to-noise reliability of 0.968 for treatment initiation and a median signal-to-noise reliability of 0.901 for treatment engagement.

Simulated decile tables:

Computer simulation was used to create datasets for treatment initiation and engagement that mirrors, as closely as possible, the mean, standard deviation, and percentile information provided for the performance score and calculated reliability. Tables 3.16.3, 3.16.4, 3.16.5, and 3.16.6 are created from the simulated datasets and provide reviewers with a more standardized format to assess reliability.

For Tables 3.16.3 and 3.16.4, entities were sorted by performance score, and the average score by decile (estimated from the simulated data) is reported along with the number of entities included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.16.3. MUC2023-179 Importance – Initiation (Decile by performance score)

MUC2023-179	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	33.18 (12.98)	1.14	12.80	19.30	23.53	27.08	30.47	34.02	38.03	42.13	47.23	57.34	81.15
Entities	3998	1	400	400	400	400	400	400	400	400	399	399	1

Table 3.16.4. MUC2023-179 Importance – Engagement (Decile by performance score)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	5.06 (3.73)	0.00	0.59	1.50	2.24	3.02	3.82	4.71	5.71	7.02	8.91	13.11	26.15
Entities	3998	64	400	400	400	400	400	400	400	400	399	399	1

For Tables 3.16.5 and 3.16.6, entities were sorted by reliability and the average reliability by decile (estimated from the simulated data) is reported along with the number of entities included in each average. Average, minimum, and maximum reliability and expected events are also included.

Table 3.16.5. MUC2023-179 Reliability – Initiation (Decile by reliability)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Reliability	0.93	0.78	0.87	0.90	0.92	0.93	0.93	0.94	0.95	0.96	0.96	0.97	1.00
Entities	3998	1	400	400	400	400	400	400	400	400	399	399	1

Table 3.16.6. MUC2023-179 Reliability – Engagement (Decile by reliability)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Reliability	0.85	0.57	0.69	0.76	0.80	0.82	0.85	0.87	0.90	0.92	0.95	0.98	1.00
Entities	3998	1	400	400	400	400	400	400	400	400	399	399	64

Assumptions:

The measure report estimates an annual denominator size of 868,407 for treatment initiation and engagement. A 2022 KFF report estimates there are 3,998 different Medicare Advantage plans.

Interpretation:

The reported median reliability for treatment initiation is 0.968. The reliability simulation suggests that few, if any, of the entities have reliability below 0.6.

The reported median reliability for treatment engagement is 0.901. The reliability simulation suggests that few, if any, of the entities have reliability below 0.6.

3.17 MUC2023-212 Level I Denials Upheld Rate Measure

Description: This rating shows how often a Medicare Advantage Organization review found their original determination decision to deny coverage to be reasonable.

Measure Type: Process

Level of Analysis: Health Plan

Data Source(s): Administrative Data (non-claims); Claims Data

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

The Federation of American Hospitals submitted the Level I Denials Upheld Rate measure for CMS's consideration to include in the Part C and D Star Ratings program. This measure would focus on how often a Medicare Advantage Organization review found their original determination decision to deny coverage to be reasonable.

The existing Part C Star Ratings measure called Reviewing Appeals Decisions focuses on Level 2 appeals reviewed by the Independent Review Entity. The submitted measure by the Federation of American Hospitals potentially could complement the existing Star Ratings measure by focusing on Level 1 appeals that occur earlier in the process to ensure enrollees are able to access the care they need.

We are interested in hearing feedback on this measure.

Table 3.17.1. MUC2023-212 Brief Summary of Measure Information

MERIT Submission Information MUC2023-212	Description
Measure name	Level I Denials Upheld Rate Measure
MUC ID	MUC2023-212
Cascade priority	Person-Centered Care
Measure steward	Federation of American Hospitals
Measure Developer	Federation of American Hospitals

MERIT Submission Information MUC2023-212	Description
Program submitted to	Medicare Part C & D Star Rating
Committee assigned to	Clinician Committee
Related measures in the program	Stars Ratings: C27: Reviewing Appeals Decisions The existing Stars Rating measure (Reviewing Appeals Decisions) assesses Level 2 Appeals reviewed by an external independent reviewer. The Level 1 Upheld Denial Rate measure evaluates Level 1 appeals reviewed by the health plans internally. Holding plans accountable for appropriate decision-making earlier in the appeals process will reduce the time and resource burden of external reviewers and aim to make healthcare decisions more efficient to ensure that patients get the necessary care in a timely and appropriate manner. The Level 1 Upheld Denial Rate measure addresses Level 1 appeals rather than Level 2 appeals.
Is this a new measure in this year's MUC List?	Submitted previously but not included in MUC List
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program	N/A
If previously used, please describe the history of the measure in CMS program	N/A
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	N/A
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
History of endorsement	N/A
Path to endorsement	N/A
Measure Specification Details	
Measure Description	This rating shows how often a Medicare Advantage Organization review found their original determination decision to deny coverage to be reasonable. Percent of Level 1 appeals where a plan's determination decision was "upheld" by the plan out of all the reconsiderations made by a plan (upheld, overturned, and partially overturned determinations). This is calculated as:

MERIT Submission Information MUC2023-212	Description
	([Determinations Upheld] / ([Determinations Upheld] + [Determinations Overturned] + [Determinations Partially Overturned]))* 100.
Data source	Administrative Data (non-claims); Claims Data
Level of analysis	Health Plan
Numerator	Total number of Part C Level 1 reconsiderations where a plan's decision to deny the original claim was "upheld" by the plan.
Denominator	All Level 1 reconsiderations decisions (upheld, overturned, partially overturned) that the plan made.
Numerator exceptions	N/A
Denominator exclusions	If the minimum number of appeals (upheld + overturned + partially overturned) is = 10, the result is "Not enough data available." Dismissed and Withdrawn appeals are excluded from this measure.
Denominator exceptions	N/A
Risk adjustment	No
Development Status	Fully Developed
If not fully developed, development stage	N/A
Target population	Medicare Advantage
Measure type	Process
Is the measure a composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.17.2. MUC2023-212 Measure Evaluation

MUC2023-212 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	Two reports from the Office of the Inspector General found widespread and persistent issues with claim denials in Medicare Advantage Organizations. Delays or denials of necessary care due to incorrect denials leads to treatment abandonment, negatively impacts patient outcomes, and increases unnecessary provider burden.[100, 101] Organization Determinations and Reconsiderations data are already required to be submitted by Medicare Advantage Organizations to CMS.	--	While the study population differs from the target quality program population, the importance for the selected program population can be extrapolated.
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	The measure includes all Medicare Advantage Organization contracts required to submit Organization Determinations and Reconsiderations/Redeterminations data to CMS that meet minimum enrollment criteria.	The measure developer did not submit empirical evidence of conformance.	Most persons and entities in the quality program population are included in the specification.

[100] U.S. Department of Health and Human Services Office of Inspector General. (2018). Medicare Advantage Appeal Outcomes and Audit Findings Raise Concerns About Service and Payment Denials. Accessed May 2, 2023.
[101] U.S. Department of Health and Human Services Office of Inspector General. (2022). Some Medicare Advantage Organization Denials of Prior Authorization Requests Raise Concerns About Beneficiary Access to Medically Necessary Care. Accessed May 2, 2023.

MUC2023-212 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	The measure is calculated from Organization Determinations and Reconsiderations data Medicare Advantage Organizations are required to submit to CMS.	No explicit articulation of people, processes, or technology required for measurement was submitted.	Some entities in the quality program population have access to the people, processes, and technology needed for data collection and reporting.
Importance: Do the benefits of performance improvement to the achievable benchmark of care exceed the burden to data collection and reporting? (*Context of Use)*	Generally, if plans are making appropriate decisions at the Level 1 phase of the appeals process, higher performance on Level 1 Upheld Denial Rate would be expected. Low performance may indicate that a plan is inappropriately denying care, which can lead to delayed or missed care. Most contracts have room for improvement as most uphold fewer than 50% of their denials at the Level 1 phase of the appeals process, suggesting potential unnecessary delays in or missed care. Moreover, the maximum performance was 86.23%, suggesting that contracts can perform at a higher rate. Organization Determinations and Reconsiderations data are already required to be submitted by Medicare Advantage Organizations to CMS.	--	Most of the performance improvements to the benchmark have a significant impact on quality program population outcomes.
Reliability: Is measure performance scientifically sound?	Signal-to-noise reliability testing for 536 Medicare Advantage contracts resulted in a mean reliability of 0.94,	--	Most or all entities have reliability above the threshold (0.60) *within* the quality program population.

MUC2023-212 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
(Context of Use)	indicating high reliability. Using the RAND Corporation's 0.9 mean reliability threshold, this measure can be reliably used to draw accurate conclusions about Medicare Advantage contract performance on the proposed Level 1 Upheld Denial Rate measure. Distribution of results: Min=0.48 10th=0.79 Median=0.98 90th=1.00 Max=1.00 Based on existing scientific evidence on the different interpretations and methods of estimating reliability, CMS finalized in the CY 2022 Physician Fee Schedule (86 FR 64996) rule that the 0.4 threshold for mean reliability continues to be appropriate for indicating moderate reliability for performance measures in the MIPS Cost category. Mean reliability levels above 0.7 continue to demonstrate high reliability for cost measures, as previously established in the CY 2017 Quality Payment Program final rule (81 FR 77169 through 77171). The median reliability based on the simulated dataset is 0.98, same as is given in the measure report, with about 94% of entities having a reliability greater than 0.7.		

MUC2023-212 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (Suitability for Selected Quality Program & Population)
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	Empirical testing demonstrated a positive relationship between the Level 1 Upheld Denial Rates and the proportion of (1) diabetic patients with blood sugar controlled, (2) controlled high blood pressure, (3) getting needed care, (4) getting appointments quickly, and (5) ratings of health plan. There was a negative relationship between Level 1 Upheld Denial Rates and (1) plan all-cause readmissions and (2) members choosing to leave plan.	The materials submitted did not include an explicit articulation of the way an entity may improve performance on the measure focus.	There is an association between the entity and the measure focus *within* the quality program population.
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	--	The measure is not risk adjusted.	N/A
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	Opportunities for improvement on this measure in the appropriate care setting are outlined in submission materials and supported through evidence of need.	--	There is an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

MUC2023-212 Measure Reliability

The performance score is a percentage of claim denials upheld for each entity.

Reliability (signal-to-noise) is calculated by $\frac{\sigma^2_{between}}{\sigma^2_{between}+\sigma^2_{within}}$. $\sigma^2_{between}$ is estimated by the variance of the performance score across the entities. σ^2_{within} is the variance (standard deviation squared) of the score within a single entity. The measure report indicates a median signal-to-noise reliability of 0.98.

Simulated decile tables:

Computer simulation was used to create a dataset that mirrors, as closely as possible, the mean, standard deviation and percentile information provided for the performance score and calculated reliability. Tables 1 and 2 are created from the simulated dataset and provide reviewers with a more standardized format to assess reliability.

For Table 3.17.3, entities were sorted by performance score, and the average score by decile (estimated from the simulated data) is reported along with the number of entities included in each average. Average, standard deviation, and minimum and maximum scores are also included.

Table 3.17.3. MUC2023-212 Importance (Decile by performance score)

MUC2023-212	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Score	22.24 (15.88)	0.00	1.99	6.10	9.85	13.33	17.09	21.35	25.92	31.99	39.74	55.18	91.36
Entities	3998	90	400	400	400	400	400	400	400	400	399	399	1

For Table 3.17.3, entities were sorted by reliability and the average reliability by decile (estimated from the simulated data) is reported along with the number of entities included in each average. Average, minimum, and maximum reliability and expected events are also included.

Table 3.17.4. MUC2023-212 Reliability (Decile by reliability)

MUC2023-212	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Mean Reliability	0.93	0.48	0.76	0.86	0.90	0.92	0.94	0.95	0.96	0.97	0.98	0.99	1.00
Entities	3998	1	400	400	400	400	400	400	400	400	399	399	90

Assumptions:

The study includes 536 entities but does not report the total number of patients included. The measure report estimates an annual denominator size of 525,714 or that many total claims. A 2022 KFF report estimates there are 3,998 different Medicare Advantage plans. For the simulation, the estimated annual denominator size, i.e., total number of patients, is 525,714 across an estimated 3,998 entities.

Interpretation:

The reported median reliability is 0.98. The reliability simulation suggests that few, if any, of the entities have reliability below 0.6.

3.18 MUC2023-199 Connection to Community Service Provider

Description: Percent of patients 18 years of age or older who screen positive for one or more of the following health-related social needs (HRSNs): food insecurity, housing instability, transportation problems, utility help needs, or interpersonal safety; and had contact with a Community Service Provider (CSP) for at least one of their HRSNs within 60 days after discharge.

Measure Type: Process

Level of Analysis: Facility

Data Source(s): Administrative Data (non-claims); Claims Data; Electronic Clinical Data (non-EHR); Electronic Health Record; Standardized Patient Assessments; Patient Reported Data and Surveys: Patient-reported data and standardized social needs assessments are used to determine patients matching the denominator of screening for HRSNs and a positive result for at least one HRSN.

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the measure Connection to Community Service Provider to the Hospital Inpatient Quality Reporting (IQR) Program measure set.

This measure was previously submitted by the measure developer, OCHIN, to the 2022 MUC List for consideration in the Merit-Based Incentive Payment System (MIPS) program (MUC2022-09) and will be implemented in the CY 2024 Physician Fee Schedule Final Rule. It has since been refined and resubmitted for consideration in the Hospital IQR program to encourage actionable steps to address patients' identified health-related social needs (HRSNs).

CMS is considering adding this measure to the HIQR program in support of an agency-wide strategic vision to achieve equity across the healthcare system. CMS has specifically prioritized the identification of key drivers of health, such as HRSNs, as critical to improving healthcare quality. Despite recent adoption of two drivers of health measures in the Hospital IQR Program, however, capturing systematic referral to community service providers to address patients' unmet HRSNs is a persistent measurement gap in the program. Thus, the proposed action measure – in tandem with the complementary Resolution of At Least 1 Health Related Social Need measure – would build upon existing quality measurement strategies to further a facility's understanding of populations served and, in turn, its focus on connecting patients with more holistic care and/or resources.

Table 3.18.1 MUC2023-199 Brief Summary of Measure Information

MERIT Submission Information MUC2023-199	Description
Measure name	Connection to Community Service Provider
MUC ID	MUC2023-199
Cascade priority	Equity
Measure steward	OCHIN
Measure developer	OCHIN
Program submitted to	Hospital Inpatient Quality Reporting Program; Medicare Shared Savings Program
Committee assigned to	Hospital and Clinician Committee
Related measures in the program	N/A
Is this a new measure in this year's MUC List?	Measure currently used in a CMS program being submitted as-is for a new or different program
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program	Yes
If previously used, please describe the history of the measure in CMS program	This measure was used in the CMMI Accountable Health Communities Pilot from 2017-2022 and in MIPS, in which it was recommended for rulemaking. This measure is currently in use in the Merit-based Incentive Payment System-Quality Program.
Any other program the measure is in use	Merit-based Incentive Payment System-Quality; Accountable Health Communities Pilot
Is this measure being proposed to meet a statutory requirement?	N/A
CBE endorsement status	Not Endorsed
CBE endorsement number if applicable	N/A
History of endorsement	N/A
Path to endorsement	N/A
Measure Specification Details	
Measure description	Percent of patients 18 years of age or older who screen positive for one or more of the following health-related social needs (HRSNs): food

MERIT Submission Information MUC2023-199	Description
	insecurity, housing instability, transportation problems, utility help needs, or interpersonal safety; and had contact with a Community Service Provider (CSP) for at least one of their HRSNs within 60 days after discharge.
Data source	Administrative Data (non-claims); Claims Data; Electronic Clinical Data (non-EHR); Electronic Health Record; Standardized Patient Assessments; Patient-Reported Data and Surveys: Patient-reported data and standardized social needs assessments are used to determine patients matching the denominator of screening for HRSNs and a positive result for at least one HRSN.
Level of analysis	Facility
Numerator	Number of patients 18 or older at time of admission who had contact with a Community Service Provider (defined as any independent, for-profit, non-profit, state, territorial, or local agency capable of addressing core or supplemental health-related social needs) for at least one of their HRSNs within 60 days after discharge.
Denominator	Number of patients admitted to the hospital who are 18 or older at time of admission who screened positive for one or more of the 5 core domains during the period of performance (quarterly).
Numerator exclusions	N/A
Denominator exclusions	• Patients who opt out of connection with Community Service Provider • Patients lost to follow-up after discharge
Denominator exceptions	N/A
Risk adjustment	No
Development status	Fully Developed
If not fully developed, development stage	N/A
Target population	All payer
Measure type	Process
Is the measure a composite or component of a composite?	No
Digital Measure Information	
Is this measure an eCQM?	No

MERIT Submission Information MUC2023-199	Description
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.18.2. MUC2023-199 Connection to Community Service Provider Measure Evaluation

MUC2023-199 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	CMMI's Comprehensive Primary Care Plus (CPC+) model reported in 2020 that 86% of ~1,500 Track 1 practices and 99% of ~1,500 Track 2 practices (together serving ~2.4M beneficiaries) are implementing DOH screening. Using a standard, validated screening tool, the CMS Accountable Health Communities (AHC) program has screened 1 million patients for HRSN in 21 states -- nearly 1/3 in hospital settings -- with 33% of beneficiaries screened having at least one HRSN. Of patients with at least one HRSN who were eligible for navigation, 74% of patients accepted navigation related to their HRSN, and 18% of patients accepting navigation either reported at least one HRSN resolved (14%) or connection with a CSP without resolution (4%).[102]	A list of citations is provided but submission does not summarize the literature cited.	While the study population differs from the target quality program population, the importance for the selected program population can be extrapolated.
Conformance: Does the measure as specified align with the conceptual intent?	When adjusting for bias and prevalence, agreement between the Accountable Health Communities (AHC) and Your Current Life Situation (YCLS) items was substantial or	Excluded persons: who opt out of connection with Community Service Provider or persons lost to follow-up after discharge.	Most persons and entities in the quality program population are included in the specification.

[102] CMMI. CPC Evaluation Annual Report. https://innovation.cms.gov/data-and-reports/2020/cpc-evaluation-annual-rep ort-2

MUC2023-199 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
(Concept of Interest)	higher (kappas > 0.60) for all social risks except housing quality (kappa = 0.52). The YCLS and Children's Health Watch (CHW) had substantial agreement (kappa 0.75) on housing.[103] Sensitivity of each two-item combination was high for the US population and high-risk demographic groups compared with the 18-item US Department of Agriculture's Core Food Security Module (CFSM). Sensitivity ranged from 96.4% for items 2 and 3 for households with children and incomes <200% of the federal poverty line, to 99.8% for items 1 and 3 for Spanish-speaking households. (Results for all combinations are available from the corresponding author upon request.) Specificity was lower, ranging from 73.7% for items 1 and 2 for households with children and incomes <100% of the federal poverty line, to 94.5% for items 2 and 3 for households with a respondent aged >60 years. Accuracy was high for all two-item combinations.[104] A Community Service Provider (CSP) is defined as any independent, for-profit, non-profit, state, territorial, or local agency	Some variability in the screening for health-related social need attributable to the selection of instrument (CBE). Multiple low-cost, low-literacy tools are available for social risk screening in clinical settings, but psychometric data are very limited. More research is needed on clinic-based screening tool reliability and validity as these factors should influence both adoption and utility.[105] Lack of specificity on what counts as a CSP.	Data element reliability and validity extrapolate to the quality program population.

[103] Lewis, C. C., Wellman, R., Jones, S. M., Walsh-Bailey, C., Thompson, E., Derus, A., ... & Sharp, A. L. (2020). Comparing the performance of two social risk screening tools in a vulnerable subpopulation. *Journal of family medicine and primary care*, *9*(9), 5026.
[104] Gundersen, C., Engelhard, E. E., Crumbaugh, A. S., & Seligman, H. K. (2017). Brief assessment of food insecurity accurately identifies high-risk US adults. *Public health nutrition*, *20*(8), 1367-1371.
[105] Henrikson, N. B., Blasi, P. R., Dorsey, C. N., Mettert, K. D., Nguyen, M. B., Walsh-Bailey, C., ... & Lewis, C. C. (2019). Psychometric and pragmatic properties of social risk screening tools: a systematic review. *American journal of preventive medicine*, *57*(6), S13-S24.

MUC2023-199 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	capable of addressing core or supplemental health-related social needs.		
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	Data to be reported through a Clinical Quality Measure (CQM) Registry (details unspecified). Data collection and reporting requires modification to workflow (details unspecified).	Patient-reported data and standardized assessments are used to determine patients matching the denominator of screening for HRSNs and a positive result for at least one HRSNs. EHR-and non-EHR electronic clinical data, as well as patient reported data, will be used to determine whether contact was made with a CSP. Administrative data will be used for measure stratification and ongoing performance monitoring (details unspecified).	Unable to determine if the people, processes, and technology required for data collection and reporting extrapolate to the quality program population. Unable to determine if the entities in the quality program population have access to people, processes, and technology needed for data collection and reporting.
Importance: Will performance improvement to the benchmark have a significant impact on population outcomes? (*Context of Use)*	Using a standard, validated screening tool, the CMS Accountable Healthcare Communities program has screened 1 million patients for HRSN in 21 states -- nearly 1/3 in hospital settings -- with 33% of	No empirical evidence that the benefits exceed the burden.	Unable to determine if the benefits of performance improvement to the benchmark have a significant impact on quality program population outcomes.

MUC2023-199 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	beneficiaries screened having at least one HRSN.[106] A reported social risk on the Accountable Health Communities (AHC) and Your Current Life Situation (YCLS) measures was strongly associated with having fair or poor self-rated health.[107] Household Food Security Survey (HFSS) questions 1 and 2 were most frequently endorsed among food-insecure families (92.5% and 81.9%, respectively). An affirmative response to either question 1 or 2 had a sensitivity of 97% and specificity of 83% and was associated with increased risk of reported poor/fair child health (adjusted odds ratio [aOR]: 1.56; P < 0.001), hospitalizations in their lifetime (aOR: 1.17; P < 0.001), and developmental risk (aOR: 1.60; P < 0.001).[108] 2441 of 3162 patients and/or caregivers who responded to the question asking whether information from the measure (e.g., the measured outcome or process) is important to know about AND can help improve care for patients in similar situations or with similar.		

[106] CMMI. CPC Evaluation Annual Report. https://innovation.cms.gov/data-and-reports/2020/cpc-evaluation-annual-rep ort-2
[107] Lewis, C. C., Wellman, R., Jones, S. M., Walsh-Bailey, C., Thompson, E., Derus, A., ... & Sharp, A. L. (2020). Comparing the performance of two social risk screening tools in a vulnerable subpopulation. *Journal of family medicine and primary care*, *9*(9), 5026.
[108] Hager, E. R., Quigg, A. M., Black, M. M., Coleman, S. M., Heeren, T., Rose-Jacobs, R., ... & Frank, D. A. (2010). Development and validity of a 2-item screen to identify families at risk for food insecurity. *Pediatrics*, *126*(1), e26-e32.

MUC2023-199 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Reliability: Is measure performance scientifically sound? *(Context of Use)*	--	Entity-level reliability not reported.	Unable to determine if entities have reliability above the threshold (0.60) *within* the quality program population.
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	Guideline: The USPSTF provides a "B" recommendation that clinicians screen for Intimate Partner Violence (one of the HRSNs included in the denominator of the proposed measure) in women of reproductive age and provide or refer women who screen positive to ongoing support services. (Note: an update on this topic is in progress—last update April 19, 2023.) USPSTF recently released a technical brief on screening and interventions for social risk factors, which notes that social risk factors are mentioned in two-thirds of USPSTF recommendation statements, and six other professional medical organizations explicitly promote clinician engagement in social risk screening and referrals.[109]	No explicit articulation of the way an entity may improve performance on the measure focus.	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population. There is no articulation of the way an entity may improve performance on the measure focus within the quality program population.
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	Strong recommendation to stratify the measures by race/ethnicity. Data from the AHC found racial/ethnic minorities were over-represented in the navigation-eligible groups. CMS has stated in its strategic plan that the imperative to stratify by race/ethnicity is a	No explicit rationale for confounders included in the model.	N/A

[109] Eder, M., Henninger, M., Durbin, S., Iacocca, M. O., Martin, A., Gottlieb, L. M., & Lin, J. S. (2021). Screening and interventions for social risk factors: technical brief to support the US Preventive Services Task Force. *JAMA*, *326*(14), 1416-1428.

MUC2023-199 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	global issue for the Agency that applies to all measures.[110]		
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	7 of 8 measured entities (or others) responded when asked if information produced by the performance measure is easy to understand AND useful for decision-making. The USPSTF report has notably highlighted the **lack** of unintended consequences encountered during implementation of social risk screening and intervention in studies reporting these outcomes, despite any perceived barriers. Availability of health information exchanges (HIEs) that facilitate the coordination between health care providers and community organizations may facilitate.	Potential for societal stigma and discrimination related to certain HRSNs (e.g., mental health issues, substance abuse) (CBE). Potential for language barriers to hinder effective communication between health care providers and patients (CBE). Potential for low degree of trust in the health care system or fear of negative consequences (e.g., immigration status concerns) (CBE). One potential unintended consequence of the measure is that hospitals might not be equipped to act on it due, in part, to the lack of community resources. This challenge was	There is not an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

[110] CMS. Health Equity Fact Sheet. https://www.cms.gov/files/document/health-equity-fact-sheet.pdf

MUC2023-199 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
		noted as a primary barrier to connecting beneficiaries to resources in the AHC Year 1 evaluation. There is a well-documented and well-tested catalog of additional tools, infrastructure, and investments that can be implemented to support practices in acting on this measure. Other considerations include: 1) Locations with limited availability of resources, such as social workers or community support programs 2) Fragmented health care system with poor coordination among providers and community organizations 3) Rural or remote areas may have limited access to social services and community resources 4) Locations with persistent economic inequality may make it difficult to fully address HRSNs	

MUC2023-199 Measure Reliability

The performance score is the ratio of the number of patients who reported contact with a Community Service Provider for a Health-Related Social Need (HRSN) to the number of patients who screened positive for one or more HRSNs.

The measure report indicates a median signal-to-noise reliability of 0.6, but it appears that this is a measure of the agreement between screening tools. Agreement with other tools may address validity but not signal-to-noise reliability.

Interpretation:

Reliability was not analyzed for this measure according to the report provided. The single value of 0.18 (reported as the mean, minimum, and maximum) is not adequate information to simulate or assess reliability for this measure.

3.19 MUC2023-210 Resolution of At Least 1 Health-Related Social Need

Description: Percent of patients 18 years or older who screen positive for one or more of the following health related social needs (HRSNs): food insecurity, housing instability, transportation problems, utility help needs, or interpersonal safety; and report that at least 1 of their HRSNs was resolved within 12 months after discharge.

Measure Type: Outcome

Level of Analysis: Facility

Data Source(s): Administrative Data (non-claims); Claims Data; Electronic Clinical Data (non-EHR); Electronic Health Record; Standardized Patient Assessments; Patient-reported data and standardized social needs assessments are used to determine patients matching the denominator of screening for HRSNs and a positive result for at least one HRSN.; Patient-Reported Data and Surveys.

Development Status: Fully Developed

Endorsement Status: Not Endorsed

CMS-Provided Rationale for Measure Consideration:

CMS is considering adding the measure Resolution of At Least 1 Health Related Social Need to the Hospital Inpatient Quality Reporting (IQR) Program measure set.

This measure was submitted to the 2022 MUC List (2022 MUC 111) for consideration of inclusion in the Merit-Based Incentive Payment System (MIPS) program, however the measure was not finalized in any program at this time. This measure is now being submitted by the measure developer, OCHIN, for consideration in the Hospital IQR Program.

CMS is considering adding this measure to the HIQR program in support of an agency-wide strategic vision to achieve equity across the healthcare system. CMS has specifically prioritized the identification of key drivers of health, such as HRSNs, as critical to improving healthcare quality. Despite recent adoption of two drivers of health measures in the Hospital IQR Program, however, capturing eventual resolution of patients' unmet HRSNs is a persistent measurement gap in the program. Thus, the proposed action measure – in tandem with the complementary Connection to Community Service Provider measure – would build upon existing quality measurement strategies to further a facility's understanding of populations served and, in turn, its focus on meaningfully and holistically addressing patient needs.

Table 3.19.1. MUC2023-210 Brief Summary of Measure Information

MERIT Submission Information MUC2023-210	Description
Measure name	Resolution of At Least 1 Health-Related Social Need
MUC ID	MUC2023-210
Cascade priority	Equity
Measure steward	OCHIN
Measure Developer	OCHIN
Program submitted to	Hospital Inpatient Quality Reporting Program; Medicare Shared Savings Program
Committee assigned to	Hospital and Clinician Committee
Related measures in the program	N/A
Is this a new measure in this year's MUC List?	Yes
If not a new measure, then describe the history of this measure in prior MUC list inclusion	N/A
Is the measure currently used in a CMS program	Merit-based Incentive Payment System-Quality; Accountable Health Communities Pilot
If previously used, please describe the history of the measure in CMS program	Measure currently used in a CMS program being submitted as-is for a new or different program. CMMI Accountable Health Communities Pilot (2017-2022); MIPS (recommended for rulemaking)
Any other program the measure is in use	N/A
Is this measure being proposed to meet a statutory requirement?	N/A
CBE endorsement status	N/A
CBE endorsement number if applicable	N/A
History of endorsement	Not Endorsed
Path to endorsement	Unknown
Measure Specification Details	
Measure Description	Percent of patients 18 years or older who screen positive for 1 or more of the following health related social needs (HRSNs): food insecurity,

MERIT Submission Information MUC2023-210	Description
	housing instability, transportation problems, utility help needs, or interpersonal safety; and report that at least 1 of their HRSNs was resolved within 12 months after discharge.
Data source	Administrative Data (non-claims); Claims Data; Electronic Clinical Data (non-EHR); Electronic Health Record; Standardized Patient Assessments; Patient-reported data and standardized social needs assessments are used to determine patients matching the denominator of screening for HRSNs and a positive result for at least one HRSN.; Patient-Reported Data and Surveys. EHR-and non-EHR electronic clinical data, as well as patient reported data, will be used to determine whether contact was made with a CSP. Administrative data will be used for measure stratification and ongoing performance monitoring.
Level of analysis	Facility
Numerator	Number of patients 18 or older at time of admission who report that at least one of their HRSNs was resolved within 12 months after discharge (quarterly).
Denominator	Number of patients admitted to the hospital who are 18 or older at time of admission who screened positive for one or more of the 5 core domains in the 12 months prior to the period of performance (quarterly).
Numerator exclusions	N/A
Denominator exclusions	• Patients who opt out of connection with Community Service Provider • Patients lost to follow-up after discharge
Denominator exceptions	N/A
Risk adjustment	No
Development Status	Fully Developed
If not fully developed, development stage	N/A
Target population	All payer
Measure type	Outcome
Is the measure a composite or component of a composite?	No

MERIT Submission Information MUC2023-210	Description
Digital Measure Information	
Is this measure an eCQM?	No
If eCQM, what is the Measure Authoring Tool (MAT) number?	N/A
If eCQM, does the measure have a Health Quality Measures Format (HQMF) specification in alignment with the latest HQMF and eCQM standards, and does the measure align with Clinical Quality Language (CQL) and Quality Data Model (QDM)?	N/A

Table 3.19.2. MUC2023-210 Resolution of At Least 1 Health-Related Social Need Measure Evaluation

MUC2023-210 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Importance: Does the measure align with goals and priorities? (*Concept of Interest)*	CMMI's Comprehensive Primary Care Plus (CPC+) model reported in 2020 that 86% of ~1,500 Track 1 practices and 99% of ~1,500 Track 2 practices (together serving ~2.4M beneficiaries) are implementing DOH screening. Using a standard, validated screening tool, the CMS Accountable Health Communities (AHC) program has screened 1 million patients for HRSN in 21 states -- nearly 1/3 in hospital settings -- with 33% of beneficiaries screened having at least one HRSN. Of patients with at least one HRSN who were eligible for navigation, 74% of patients accepted navigation related to their HRSN, and 18% of patients accepting navigation either reported at least one HRSN resolved	--	While the study population differs from the target quality program population, the importance for the selected program population can be extrapolated.

MUC2023-210 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	(14%) or connection with a CSP without resolution (4%).[111] The measure developer summarized the relevance of the measure to the Centers for Medicare and Medicaid Services, they did not provide a summary of the literature on the benefits of measuring each of the 'five core domains' used to operationalize determinants of health.		
Conformance: Does the measure as specified align with the conceptual intent? *(Concept of Interest)*	When adjusting for bias and prevalence, agreement between the Accountable Health Communities (AHC) and Your Current Life Situation (YCLS) items was substantial or higher (kappa > 0.60) for all social risks except housing quality (kappa = 0.52). The YCLS and Children's Health Watch (CHW) had substantial agreement (kappa 0.75) on housing.[112] Sensitivity of each two-item combination was high for the US population and high-risk demographic groups compared with the eighteen-item US Department of Agriculture's Core Food Security Module (CFSM). Sensitivity ranged from 96.4% for items 2 and 3 for households with children and incomes <200% of the federal poverty line, to 99.8% for items 1 and 3 for Spanish-speaking households. Specificity was lower, ranging from 73.7% for items 1 and 2 for households with children	Excluded persons: who opt out of connection with Community Service Provider or who are lost to follow-up after discharge. Some variability in the screening for health-related social need attributable to the selection of instrument (CBE). Multiple low-cost, low-literacy tools are available for social risk screening in clinical settings, but psychometric data are very limited. More research is needed on clinic-based screening tool reliability and validity as these factors should influence both adoption and utility.	Most persons and entities in the quality program population are included in the specification. Data element reliability and validity extrapolate to the quality program population.

[111] CMMI. CPC Evaluation Annual Report. https://innovation.cms.gov/data-and-reports/2020/cpc-evaluation-annual-rep ort-2
[112] Lewis, C. C., Wellman, R., Jones, S. M., Walsh-Bailey, C., Thompson, E., Derus, A., ... & Sharp, A. L. (2020). Comparing the performance of two social risk screening tools in a vulnerable subpopulation. *Journal of family medicine and primary care*, *9*(9), 5026.

MUC2023-210 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	and incomes <100% of the federal poverty line, to 94.5% for items 2 and 3 for households with a respondent aged >60 years. Accuracy was high for all two-item combinations.[113] A Community Service Provider (CSP) is defined as any independent, for-profit, non-profit, state, territorial, or local agency capable of addressing core or supplemental health-related social needs.		
Feasibility: Does the measure's specification and data collection minimize burden? *(Concept of Interest)*	Data to be reported through a Clinical Quality Measure (CQM) Registry (details unspecified). Data collection and reporting requires modification to workflow (details unspecified).	Patient-reported data and standardized assessments are used to determine patients matching the denominator of screening for HRSNs and a positive result for at least one HRSNs. EHR-and non-EHR electronic clinical data, as well as patient reported data, will be used to determine whether contact was made with a CSP. Administrative data will be used for measure stratification and ongoing performance monitoring (details unspecified).	Unable to determine if the people, processes, and technology required for data collection and reporting extrapolate to the quality program population. Unable to determine if the entities in the quality program population have access to people, processes, and technology needed for data collection and reporting.
Importance: Will performance improvement to the benchmark have a	Using a standard, validated screening tool, the CMS Accountable Healthcare Communities program has screened 1 million patients for HRSN in 21 states -- nearly 1/3 in hospital settings -- with 33% of	Performance scores not reported. No empirical evidence that the benefits exceed the burden.	Unable to determine if the benefits of performance improvement to the benchmark have a significant impact on

[113] Gundersen, C., Engelhard, E. E., Crumbaugh, A. S., & Seligman, H. K. (2017). Brief assessment of food insecurity accurately identifies high-risk US adults. *Public health nutrition*, *20*(8), 1367-1371.

MUC2023-210 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
significant impact on population outcomes? (*Context of Use)*	beneficiaries screened having at least one HRSN.[114] A reported social risk on the Accountable Health Communities (AHC) and Your Current Life Situation (YCLS) measures was strongly associated with having fair or poor self-rated health.[115] Household Food Security Survey (HFSS) questions 1 and 2 were most frequently endorsed among food-insecure families (92.5% and 81.9%, respectively). An affirmative response to either question 1 or 2 had a sensitivity of 97% and specificity of 83% and was associated with increased risk of reported poor/fair child health (adjusted odds ratio [aOR]: 1.56; P < .001), hospitalizations in their lifetime (aOR: 1.17; P < 0.001), and developmental risk (aOR: 1.60; P < 0.001).[116] 2441 of 3162 patients and/or caregivers responded to the question asking whether information from the measure (e.g., the measured outcome or process) is important to know about AND can help improve care for patients in similar situations or with similar.		quality program population outcomes.

[114] CMMI. CPC Evaluation Annual Report. https://innovation.cms.gov/data-and-reports/2020/cpc-evaluation-annual-report-2
[115] Lewis, C. C., Wellman, R., Jones, S. M., Walsh-Bailey, C., Thompson, E., Derus, A., ... & Sharp, A. L. (2020). Comparing the performance of two social risk screening tools in a vulnerable subpopulation. *Journal of family medicine and primary care*, *9*(9), 5026.
[116] Hager, E. R., Quigg, A. M., Black, M. M., Coleman, S. M., Heeren, T., Rose-Jacobs, R., ... & Frank, D. A. (2010). Development and validity of a 2-item screen to identify families at risk for food insecurity. *Pediatrics*, *126*(1), e26-e32.

MUC2023-210 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
Reliability: Is measure performance scientifically sound? *(Context of Use)*	--	Entity-level reliability not reported.	Unable to determine if entities have reliability above the threshold (0.60) *within* the quality program population.
Validity: May providers/facilities/care systems effectively improve on this measure? *(Context of Use)*	Guideline: The USPSTF provides a "B" recommendation that clinicians screen for Intimate Partner Violence (one of the HRSNs included in the denominator of the proposed measure) in women of reproductive age and provide or refer women who screen positive to ongoing support services. (Note: an update on this topic is in progress—last update April 19, 2023.) USPSTF recently released a technical brief on screening and interventions for social risk factors which notes that social risk factors are mentioned in two-thirds of USPSTF recommendation statements, and six other professional medical organizations explicitly promote clinician engagement in social risk screening and referrals.[117]	No explicit articulation of the way an entity may improve performance on the measure focus.	There is an association between the entity and the measure focus in a population that extrapolates to the quality program population. There is no articulation of the way an entity may improve performance on the measure focus within the quality program population.
Threats to Validity: If appropriate, is the measure risk adjusted to account for factors outside entity control? *(Context of Use)*	Strong recommendation to stratify the measures by race/ethnicity. Data from the AHC found racial/ethnic minorities were over-represented in the navigation-eligible groups.[118]	No explicit rationale for confounders included in the model.	N/A

[117] Eder, M., Henninger, M., Durbin, S., Iacocca, M. O., Martin, A., Gottlieb, L. M., & Lin, J. S. (2021). Screening and interventions for social risk factors: technical brief to support the US Preventive Services Task Force. *JAMA*, *326*(14), 1416-1428.
[118] CMMI. CPC Evaluation Annual Report. https://innovation.cms.gov/data-and-reports/2020/cpc-evaluation-annual-rep ort-2

MUC2023-210 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
	CMS has stated in its strategic plan that the imperative to stratify by race/ethnicity is a global issue for the Agency that applies to all measures.[119]		
Usability: Is there opportunity for improvement on this measure in the intended use setting? *(Context of Use)*	7 of 8 (87%) measured entities (or others) responded when asked if information produced by the performance measure is easy to understand AND useful for decision-making. USPSTF report has notably highlighted the **lack** of unintended consequences encountered during implementation of social risk screening and intervention in studies reporting these outcomes, despite any perceived barriers.[120] Reasonable availability of health information exchanges (HIEs) that facilitate the coordination between health care providers and community organizations.	Potential for societal stigma and discrimination related to certain HRSNs (e.g., housing insecurity, experiences of intimate partner violence) (CBE). One potential unintended consequence of the measure is that hospitals might not be equipped to act on it due, in part, to the lack of community resources. This challenge was noted as a primary barrier to connecting beneficiaries to resources in the AHC Year 1 evaluation. There is a well-documented and well-tested catalog of additional tools, infrastructure, and investments that can be implemented to support practices in acting on this measure. Other considerations include: 1) Locations with limited availability of resources, such	There is not an explicit articulation of the resources and context that might facilitate improvement that extrapolates to the quality program population.

[119] CMS. Health Equity Fact Sheet. https://www.cms.gov/files/document/health-equity-fact-sheet.pdf
[120] Eder, M., Henninger, M., Durbin, S., Iacocca, M. O., Martin, A., Gottlieb, L. M., & Lin, J. S. (2021). Screening and interventions for social risk factors: technical brief to support the US Preventive Services Task Force. *JAMA*, *326*(14), 1416-1428.

MUC2023-210 Criteria/Assertions	Measure Benefits & Evidence Supporting Inclusion	Areas for Additional Consideration	External Validity (suitability for selected quality program and population)
		as social workers or community support programs 2) Fragmented health care system with poor coordination among providers and community organizations 3) Rural or remote areas may have limited access to social services and community resources 4) Locations with persistent economic inequality may make it difficult to fully address HRSNs.	

MUC2023-210 Measure Reliability

The performance score is ratio of the number of patients who reported at least one HRSN resolved to the number of patients who screened positive for one or more HRSNs.

The measure report indicates a median signal-to-noise reliability of 0.6, but it appears that this is a measure of the agreement between screening tools. Agreement with other tools may address validity but not signal-to-noise reliability.

Interpretation:

Reliability was not analyzed for this measure, according to the report provided. The single value of 0.18 (reported as the mean, minimum, and maximum) is not adequate information to simulate or assess reliability for this measure.

Appendix A. Excerpts from the CMS 2023 Measures Under Consideration List Program-Specific Measure Needs and Priorities[121]

Merit-based Incentive Payment System Program

Program History and Structure:

- The Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) ended the Sustainable Growth Rate (SGR) formula, which would have resulted in a significant cut to payment rates for clinicians participating in Medicare. MACRA requires CMS, by law, to implement an incentive program for clinicians. This program, referred to as the Quality Payment Program, provides two participation pathways for clinicians:
  - The Merit-based Incentive Payment System (MIPS): Traditional MIPS or MIPS Value Pathways (MVPs).
  - Advanced Alternative Payment Models (Advanced APMs.)
- MIPS combines three Medicare "legacy" programs—the Physician Quality Reporting System (PQRS), Value-based Payment Modifier (VM), and the Medicare EHR Incentive Program for Eligible Professionals—into a single program. Under MIPS, four connected performance categories will affect a clinician's future Medicare payments.
- Starting with the 2023 performance period, MIPS eligible clinicians and groups may choose to report traditional MIPS or MIPS MVPs. MVPs include a subset of measures and activities that are related to a given specialty or medical condition. MVPs offer reduced reporting requirements, allowing MVP participants to report on a smaller, more cohesive subset of measures and activities (within the measures and activities available under traditional MIPS).
- Each performance category is scored independently and has a specific weight, indicating its contribution towards the MIPS Final Score.
- The MIPS performance categories and their 2022 weights toward the final score are: Quality (30%); Promoting Interoperability (25%); Improvement Activities (15%); and Cost (30%). The final score (100%) will be the basis for the MIPS payment adjustment assessed for MIPS eligible clinicians. The following tables categorize the 2023 performance period MIPS quality measure inventory based on measure type and the Meaningful Measures 2.0 Framework Domains.

[121] CMS. 2023 MUC List Program Specific Measure Needs and Priorities. Accessed 8th November 2023. https://mmshub.cms.gov/sites/default/files/2023-MUC-List-Program-Specific-Measure-Needs-and-Priorities.pdf

Measure Type	Number of Measures (2023 Performance Period)
Patient Engagement/Experience	2
Efficiency	5
Intermediate Outcome	7
Outcome	33
Patient Reported Outcome-Based Performance Measure (PRO-PM)	17
Process	133
Structure	1
Total	**198**

Meaningful Measures 2.0 Priority	Number of Measures (2023 Performance Period)
Person-centered Care	33
Equity	1
Safety	37
Affordability and Efficiency	28
Chronic Conditions	44
Wellness and Prevention	23
Seamless Care Coordination	10
Behavioral Health	22
Total	**198**

**MIPS Meaningful Measures counts may shift as categorizations are finalized for 9 measures new in 2023.*

Part C and D Star Ratings

Program History and Structure:

- The Part C & D Star Ratings program is based on sections 1851(d), 1852(e), 1853(o) and the 1854(b)(3)(iii), (v), and (vi) of the Social Security Act.
- General authority under section 1856(b) of the Act: establishment of standards consistent with and to carry out Part C & D as basis for the 5-Star Ratings system.
- The methodology for the Part C & D Star Ratings program was codified in contract year (CY) 2019 Medicare Part C and D Final Rule.
- CMS must propose through rulemaking any changes to the methodology for calculating the Star Ratings, the addition of new measures, the removal of a measure within the Star Ratings, and substantive measure changes per §423.184 and §422.164.
- Non-substantive measure specification changes for the Star Ratings will be announced through the advance notice process per §423.184(d)(1) and §422.164(d)(1).
- The Star Ratings Program is consistent with CMS's Quality Strategy of optimizing health outcomes by improving quality and transforming the health care system. The CMS Quality Strategy goals reflect the six priorities set out in the National Quality Strategy: safety, person and caregiver-centered experience and outcomes, care coordination, clinical care, population/community health, efficiency, and cost reduction.
    - CMS highlights contracts receiving an overall rating of 5 stars with the High Performing Icon (HPI) on the MPF:
- Beneficiaries may enroll in a 5-Star PDP, MA-PD, or MA-only plan through a Special Election Period (SEP). 5-Star Plans may market year-round.
- Beneficiaries may not enroll online via the MPF in a Low Performing Icon (LPI) plan. Beneficiaries must contact the plan directly.
- The LPI Icon is displayed for contracts rated less than 3 stars for at least the last 3 years in a row for their Part C or D summary rating.
- Beneficiaries in LPI plans are eligible for a Special Enrollment Period (SEP) to move to a higher quality plan.
- Per the Affordable Care Act, CMS makes Quality Bonus Payments (QBPs) to MA organizations that meet quality standards measured using a five-star quality rating.
- The QBP percentage for each Star Rating for 2020 payments:

Star Rating	QBP Percentage
3.5 stars or below	0%
4 stars or more	**5%**

- The MA rebate level for plans is tied to the contract's Star Rating.

Current Measure Information:

Measure Type	Number of Measures
Composite	0
Cost/Resource Use	0
Intermediate Outcome	5
Outcome	2
Patient Reported Outcome-Based Performance Measure (PRO-PM)	8
Process	26
Structure	0
Total	**41 (38 unique measures)**

Meaningful Measures 2.0 Priority	Number of Measures
Person-centered Care	18
Equity	0
Safety	2
Affordability and Efficiency	1
Chronic Conditions	9
Wellness and Prevention	4
Seamless Care Coordination	4
Behavioral Health	0
Total	**38**

Medicare Shared Savings Program

Program History and Structure:

- The Medicare Shared Savings Program (Shared Savings Program) is Medicare's national value-based payment program for Accountable Care Organizations (ACO). ACO's facilitate coordination and cooperation among health care providers to improve the quality of care for Medicare Fee-For-Service (FFS) beneficiaries and reduce the rate of growth in health care costs
- Eligible clinicians, hospitals, and other health care providers can voluntarily join or form an ACO
- ACOs share in savings by meeting the quality performance standard for the performance year and lowering the growth in Medicare spending
- ACOs participating under a two-sided shared savings/losses model may owe losses if they increase costs and the amount owed is based on quality performance depending on track
- For performance years 2023 and 2024, ACOs will be required to report quality data via the Alternative Payment Model (APM) Performance Pathway (APP).
- ACOs can choose to report either the 10 measures under the CMS Web Interface or the 3 eCQMs/Merit-based Incentive Payment System (MIPS) Clinical Quality Measures (CQMs)
- ACOs must field the Consumer Assessment of Healthcare Providers and Systems Survey (CAHPS) for MIPS survey
- CMS will calculate 2 claims-based outcome measures using administrative claims data: the Hospital Wide, 30-day, All-Cause Unplanned Readmission (HWR) Rate for MIPS Eligible Clinician Groups measure and the Clinician and Clinician Group Risk-Standardized Hospital Admissions Rates for Patients with Multiple Chronic Conditions measure
- For performance year 2025 and subsequent performance years, ACOs will be required to report:
- the 3 eCQMs/MIPS CQMs, field the CAHPS for MIPS survey, and CMS will continue to calculate the 2 claims-based outcome measures noted above

Current Measure Information:

Measure Type	Number of Measures
Composite	0
Cost/Resource Use	0

Measure Type	Number of Measures
Intermediate Outcome	2
Outcome	3
Patient Reported Outcome-Based Performance Measure (PRO-PM)	1
Process	7
Structure	0
Total	**13**

Meaningful Measures 2.0 Priority	Number of Measures
Person-centered Care	1
Equity	0
Safety	1
Affordability and Efficiency	1
Chronic Conditions	4
Wellness and Prevention	4
Seamless Care Coordination	0
Behavioral Health	2
Total	**13**