2024 Pre-Rulemaking Measure Review Preliminary Assessment

MUC ID	Title
MUC2024-049	Breast Cancer Screening
Measure Steward & Developer	**Proposed CMS Programs**
Acumen, LLC	Merit-based Incentive Payment System-Cost

Measure Overview
Developer-provided rationale (excerpt from submission): The Breast Cancer Screening episode-based cost measure was selected for development because it fills a measurement and performance gap in MIPS, such as a lack of episode-based cost measures applicable to radiologists. This measure addresses unmet needs in MIPS, as none of the current clinician-level measures evaluate the outcome of screening mammograms and goes beyond avoiding diagnostic errors to account for cost and patient experience.
CMS-provided program rationale: CMS is considering adding the Breast Cancer Screening episode-based cost measure to the MIPS measure set. This cost measure was developed with input from a technical expert panel (TEP) composed of clinical experts with expertise directly relevant to breast cancer, as well as patients, caregivers, and patient advocates with lived experience of breast cancer screening, diagnosis, and treatment. This measure focuses on the clinical topic of breast cancer screening and fills measurement gaps among radiologists, a specialty with no existing cost measures in MIPS that has been identified by CMS as a high-priority area. Additionally, literature and testing suggest that there are opportunities for improvement in performance across clinicians. Finally, when paired with relevant breast cancer screening quality measures, the Breast Cancer Screening measure can be used in a potential value pathway to reward diagnostic excellence.
Description: The Breast Cancer Screening episode-based cost measure evaluates a clinician's or clinician group's average risk-adjusted cost to Medicare for providing care to females 40 years of age or older, who received a screening mammogram during an episode of care. The measure score is the clinician's or clinician group's risk-adjusted cost for the episode group averaged across all episodes attributed to the clinician or clinician group. This measure includes costs for certain services (calculated from claims submitted under Medicare Parts A and B) that are clinically related to the attributed clinician's or clinician group's role in managing care during each episode. Each episode starts from the screening mammogram that opens, or "triggers," the episode and continues through 12 months after the trigger or the next screening mammogram. This measure would assess the costs of certain assigned services clinically related to breast cancer screening, including basic and advanced diagnostic services and cancer treatment services.

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.

Measure Overview	
Measure background: New measure, never reviewed by Measure Applications Partnership (MAP) Workgroup or Pre-Rulemaking Measure Review (PRMR) or used in a Medicare program.	
Numerator: The cost measure numerator is the sum of the ratio of observed to expected cost to Medicare Parts A & B for all Breast Cancer Screening episodes attributed to a clinician or clinician group. This sum is then multiplied by the national average observed episode cost to generate a dollar figure. **Exclusions:** N/A	
Denominator: The cost measure denominator is the total number of episodes from the Breast Cancer Screening episode group attributed to a clinician. **Exclusions:** The measure denominator excludes: • Male patients • Patients under 40 years of age • Patients with a history of breast cancer • Patients not found in Master Beneficiary Summary File • Patients residing outside of the United States or its territories • Patients missing date of birth • Patients not continuously 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 • Patients with death date occurring during the episode • Patients with extremely low episode costs **Exceptions:** N/A	
Measure type: Cost/Resource Use	**Measure has multiple scores:** No **Measure is a composite:** No **Measure is digital and/or an eCQM:** No **Measure is a paired or group measure:** No
Level of analysis: Clinician: Individual and Group	**Data source(s):** Digital-Administrative systems: Claims Data
Care setting(s): Ambulatory/office-based care	**Risk adjustment or stratification:** Risk adjusted
CBE endorsement status: Never submitted	**CBE endorsement history:** None
Is measure currently used in CMS programs? No	**Measure addresses statutorily required area?** The development and implementation of cost measures in MIPS is required by Section 101(f) of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA).

Meaningfulness

Importance	
Type of evidence:	Clinical Guidelines or USPSTF (U.S. Preventive Services Task Force) Guidelines; Peer-Reviewed Systematic Review; Peer-Reviewed Original Research; Empirical data [Source: Measures Under Consideration (MUC) Entry/Review Information Tool (MERIT) Submission Form]
Importance: The American Cancer Society recommends breast cancer screening using 2D or 3D mammography for women as young as 40 years, and annually or biennially for older women while in good health (guideline is ungraded). The measure submission identified seven systematic reviews; findings included interventions shown to increase screening, mixed results related to follow-up time after positive screen and potential for overdiagnosis, higher costs associated with double-reading screening results, lower costs associated with diagnosing cancer at stage 1, and lower mortality associated with screening. The review of empirical evidence indicates that one in eight women will develop breast cancer in their lifetimes and that appropriate screening detects cancers earlier and reduces mortality and costs. The developer did not evaluate meaningfulness to patients, but their submission did note that the technical expert panel (TEP) included three persons and family partners with lived experience of breast cancer screening, diagnosis, and treatment. [Source: Evidence-Reliability-Validity-Testing]	
Rating: Met	

Measure Performance

For Tables 1a and 1b, the developer provided the mean score and number of entities for each performance score decile (i.e., the data sorted and broken into 10 equal parts) along with the minimum and maximum performance scores observed.

Interpretation: The mean scores for the 2,441 TIN (Taxpayer Identification Number, assigned by the IRS) entities and 16,289 TIN-NPI (Taxpayer Identification Number National Provider Identifier, assigned by CMS) entities described in the testing submission for this measure were $247.3 and $250.2 respectively. For this continuous variable measure, a lower score indicates better quality of care.

Table 1a. MUC2024-049 Performance Score Deciles – TIN (Taxpayer Identification Number)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Max
Mean Score (SD)	$247.3 ($43.8)	$120.9	$196.6	$216.2	$225.1	$234.6	$247.6	$259.4	$267.3	$277.3	$292.1	$587.4

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Max
Number of Entities	2,441	1	244	244	244	244	244	244	244	244	244	1

Table 1b. MUC2024-049 Performance Score Deciles – TIN-NPI (Taxpayer Identification Number- National Provider Identifier)

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Max
Mean Score (SD)	$250.2 ($40.1)	$101.3	$206.0	$219.6	$228.6	$238.6	$250.7	$260.7	$268.9	$279.0	$294.6	$586.6
Number of Entities	16,289	1	1,629	1,629	1,629	1,629	1,629	1,629	1,629	1,629	1,629	1

Conformance
Measure alignment with conceptual intent: The measure specification clearly identifies care episodes that are triggered by a screening event, and the process for attributing costs for basic diagnostic services, advanced diagnostic services, and breast cancer treatment to clinicians. An episode is defined as “12 months after the trigger event or the next screening mammogram” [Source: MIF]. Detailed specifications are provided for all services with costs assigned to an episode. Developers described how the TEP was used to help determine the target population, episode window, clinically related services, service assignment, and codes. Specification aligns with conceptual intent. Numerator and denominator populations are appropriate and exclusions align with clinical evidence.
Rating: Met

Feasibility	
eCQM Feasibility testing conducted:	No [Source: MERIT Submission Form]
Feasibility: All data elements required for the measure are defined in electronic sources and align with United States Core Data for Interoperability (USCDI)/USCDI+ quality standard definitions. Collecting the measure data has no effect on provider workflow.	
Rating: Met	

Validity	
Validity testing method(s):	Empiric Validity; Face Validity [Source: MERIT Submission Form, MIF]
Testing level(s):	Clinician – Group; Clinician – Individual [Source: MERIT Submission Form]
Validity: Developer described empirical validity testing procedures, which were performed at both the clinician group and individual clinician levels. The submission presents a mediation analysis of direct and indirect effects of treatment choices on the measure score and shows that spending on ambulatory/minor procedures, anesthesia, and laboratory testing is statistically associated with better measure scores at both levels. On the other hand, the main drivers of cost at both levels are major procedures, outpatient evaluation and management services, and imaging services. While major procedures show a statistical association with lower costs of adverse events, the reduction in costs of adverse events is not enough to offset the cost of major procedures, which further reinforces the need for early detection to avoid high-intensity treatments. Even though outpatient evaluation and management and imagining are essential, the results indicate that they can be prone to overuse because more spending on these services does not offset the costs of adverse events. Correlation analyses showed weak ($0.21 < r < 0.32$) positive relationships at both levels between the measure score and the Breast Cancer Screening Recall Rate measure and OP-39 Breast Cancer Screening Recall Rate, and very weak ($r > -0.15$) negative relationships at both levels between the measure score and the Breast Cancer Screening with an Eventual Breast Cancer Diagnosis: PPV1 and Use of Biopsy After Diagnostic Follow-up with an Eventual Breast Cancer Diagnosis: PPV3 measures. None of the measures used in the correlation analysis are endorsed, and PPV1 and PPV3 have not been evaluated for validity by the CBE [Source: MIPS Article]. In face validity testing, out of eight voting experts, six agreed and one strongly agreed the cost measure could distinguish good from poor quality care, and one was undecided. For patients for whom breast cancer is detected during an episode, only costs associated with basic diagnostic services are assigned to clinicians with “early breast cancer detection episodes, which are episodes with a breast cancer detection within eight months of a screening mammogram,” while all costs are assigned to clinicians with “late breast cancer detection episodes, which are episodes with a breast cancer detection after eight months of the screening mammogram” [Source: MIF]. These assignment	

Validity
rules are intended to incentivize early detection of breast cancer and encourage screening. It would be helpful if the developer could clarify how the timing of breast cancer detection within an episode can be defined as either “early detection” or “late detection,” without reference to the cancer stage, because cancer stage at diagnosis is the factor significantly associated with cost.
Threats to validity: The developer considered potential threats to validity and selected an appropriate risk adjustment model. This measure is risk adjusted for health status, comorbidities, disability status, and age. The measure is stratified by breast cancer diagnosis status before risk adjustment. Risk adjustment does not increase or decrease explained variance in cost of individual episodes; the TEP supported the inclusion of risk factors in model; risk model demonstrated good calibration/predictive ability across deciles. The TEP assisted in identifying relevant risk factors for adjustment. The committee could consider where it is appropriate to adjust for age and disability status without first establishing whether a disparity exists in referral for breast cancer screening for older or disabled patients.
Rating: Met

Reliability	
Reliability testing method(s):	Signal-to-Noise [Source: MERIT Submission Form, MIF]
Testing level:	Clinician – Group; Clinician – Individual
Reliability discussion: The dataset consists of 4,694,283 beneficiaries divided across 2,441 TINs (i.e., clinician groups) and 16,289 TIN-NPIs (individual clinicians). The median reliability for the TINs and TIN-NPIs is 0.989 and 0.958, respectively, at a 20-episode volume threshold. Both of these scores are >0.60, indicating that the measure is capable of differentiating entities by quality of performance. If the measure is implemented in MIPS in the future, CMS will establish a case minimum through notice-and-comment rulemaking.	
Additional reliability analyses: Only a single estimate for reliability is required; therefore, interpolated decile averages of the reliability data were not generated.	
Rating: Met	

Usability	
Usability considered in application:	Yes
Usability discussion: Based on discussion of the measure in the MUC List submission documents, there is an opportunity for improvement on the measure target among clinician and clinician groups participating in the MIPS program. No external program-level factors that may present barriers to measure use were identified during review. The submission does not specify any unintended consequences, but other parts of the submission suggest that increased costs associated with over-screening and overuse of diagnostic services is a possibility. This measure is not currently in use in a Medicare program.	
Rating: Met	

External Validity	
Was this measure tested in the same target population as the CMS program?	Yes
External validity discussion: The developer tested this measure in an ambulatory care setting with a target population of Medicare fee-for-service patients.	
Rating: Met	

Appropriateness of Scale

Similar or related measures in program(s):	The developer did not identify any related or competing measures.
Measure appropriateness, equity, and value across target populations/measured entities: The developer's review of active MIPS measures did not identify any similar or competing measures, suggesting that this measure would fill a gap within the current program measure set. The focus and target population of this measure align with the intent and population of the program. Regarding equity of this measure's performance and benefit across populations, the developer's literature review and sub-analysis provided does not suggest differential benefit or harm to specific subgroups of MIPS-participating clinicians or their patients. The committee should consider if, based on their professional and patient experience, there is a chance for variation on distribution of benefit or burden across provider and patient populations.	

Time to Value Realization

Plan for near- and long-term impacts after implementation:	In their evidence review and validity analyses, the developer identifies the potential for increased costs associated with over-screening and overuse of diagnostic procedures.
Measure implementation impacts over time: While the measure developer briefly mentions potential outcomes for their measure on patient populations, there may be a need for further examination of near- and long-term impacts of this measure after implementation across provider and patient populations. Questions for the committee to consider: • What are the potential near- and long-term impacts of this measure on measured entities, MIPS, and patient populations? • What are potential near- and long-term impacts of the cost measure's value proposition for health care affordability? • Will benefits and burdens associated with this measure be realized within an appropriate implementation time frame?	