Improving the Diagnostic Performance of Screening Tests for Breast Cancer

Technical Expert Panel (TEP) Meeting #3

March 2024

Acumen, LLC
500 Airport Blvd., Suite 365
Burlingame, CA 94010

PROJECT OVERVIEW

Acumen, LLC has been awarded a grant by the Gordon and Betty Moore Foundation's Diagnostic Excellence Initiative (grant number GBMF11507) to develop measures that can assess the quality and improve the value of breast cancer screening. Acumen's measure development approach involves convening a Technical Expert Panel (TEP), composed of clinical experts, as well as patients, caregivers, and patient advocates, to contribute direction and thoughtful input during measure development. The measure development content is solely the responsibility of the authors and does not necessarily represent the views of the Moore Foundation.

The application of clinically accepted metrics in breast image quality and interpretation have been associated with improved outcomes. This project will link current practice standards to policy initiatives by creating a cohesive set of measures based on the Breast Imaging Reporting and Data System (BI-RADS) Atlas®. This will allow radiology practitioners to assess performance, value, and outcomes; engage in national policy; and work towards achieving a national standard for diagnostic excellence.

Acumen is developing four clinician-level measures – three quality measures and one episode-based cost measure – to assess performance of breast imaging teams using Medicare administrative claims. Together, these could provide a cohesive set of measures for a MIPS (Merit-based Incentive Payment System) Value Pathway (MVP) to reward diagnostic excellence, currently infeasible without outcome or cost measures. Accordingly, we plan to submit the measures through the Centers for Medicare & Medicaid Services (CMS) pre-rulemaking process for consideration for use in MIPS, with the intention of them filling critical measurement gaps now present. The three quality measures and one cost measure in development that would provide more meaningful ways to assess the performance of radiologists than the measures currently available in MIPS include the following:

- Breast Cancer Screening Recall Rate Quality Measure (Outcome): Percentage of women 40 years of age and older who had a positive screening mammogram leading to timely follow-up and testing.[1]
- Breast Cancer Screening with an Eventual Breast Cancer Diagnosis: Positive Predictive Value (PPV) 1 Quality Measure (Outcome): Percentage of women 40 years of age and older who had a positive screening mammogram that led to an eventual breast cancer diagnosis.[1]

[1] Positive screening mammograms are measured as cases with diagnostic follow-up for additional testing (examples provided in later sections within this document).

- Use of Biopsy After Diagnostic Follow-up with an Eventual Breast Cancer Diagnosis: Positive Predictive Value (PPV) 3 Quality Measure (Outcome): Percentage of women 40 years of age and older who had a biopsy from a diagnostic follow-up that resulted in an eventual breast cancer diagnosis.
- Breast Cancer Screening Episode-Based Cost Measure: Costs of services that are clinically related to the attributed clinician's role in managing care during each episode, starting from the clinical event (screening mammogram) that opens, or 'triggers,' the episode through 365 days after the trigger or the next screening mammogram.

We convened a TEP to provide input on the specifications of the measures. This is part of the measure development process, as defined by the CMS Measurement Management System (MMS) Blueprint, to gather expert clinical input and individual experience of person and family partners. Acumen held a call for nominations between December 14, 2022 and January 16, 2023. Our team notified interested parties via email, and collected nominations through an online survey. A panel of 14 TEP members was finalized mid-January 2023 to provide a balanced and diverse set of perspectives. This includes clinicians with expertise directly relevant to breast cancer, as well as Person and Family Partners (PFPs), people with lived experience of breast cancer screening, diagnosis, and treatment. The TEP met for the first time virtually on January 20, 2023, and again on October 27, 2023, then met for a final time on March 8, 2024 to refine and finalize measure specifications.

TABLE OF CONTENTS

1 MEETING OVERVIEW

This meeting summary document outlines the purpose, discussion, and recommendations from the Improving the Diagnostic Performance of Screening Tests for Breast Cancer TEP meeting #3. The goals of this TEP meeting held on March 8, 2024 were the following:

(i) Gather input on the measure specifications for the three quality measures, including: a) terminology used to describe follow-up imaging; b) alignment between our Breast Cancer Screening Recall Rate measure and the Hospital Outpatient Quality Reporting (OQR) Program Breast Cancer Screening Recall Rates (OP-39) measure; c) time intervals between different services to define the measures' numerators; and d) the type of practitioners who can provide evidence of breast cancer based on their billing activity;

(ii) Gather input on the measure specifications for the episode-based cost measure, including service assignment and how to further incentivize timely and early diagnosis;

(iii) Gather input on the measure scoring methodology;

(iv) Present data on social risk factors; and

(v) Conduct a preliminary vote for face validity.

The meeting was held virtually and attended by 11 of the 14 TEP members. The meeting was facilitated by the moderator, Heather Litvinoff, and the Co-Principal Investigator, David Seidenwurm. Appendix A provides the list of TEP members and the Acumen project team. Prior to the meeting, TEP members were provided with the agenda and slides to review prior to the meeting to maximize time for discussion.

2 SUMMARY OF SESSIONS AND DISCUSSION

This section is organized based on the meeting sessions and describes TEP discussions and recommendations. Section 2.1 provides a recap of the project overview presented during the meeting. Section 2.2 provides an overview of beta testing. Session 2.3 covers the discussion on the specifications of the three quality measures. Section 2.4 summarizes the discussion on the cost measure specifications. Section 2.5 summarizes the cross-measure discussion. Section 2.6 covers the preliminary face validity vote conducted during the meeting.

2.1 Project Overview

Acumen provided a recap of the project goals and how the measures meet CMS needs and priorities. The goal of the project is to develop a set of clinical quality and cost measures on breast cancer screening and diagnosis, with the intent of submitting them to MIPS. Breast cancer screening is the focus of this project as it represents a current gap in MIPS. These measures provide a cohesive measure set that could be implemented as an MVP. Acumen has worked to ensure that the measure set containing the three quality measures and one cost measure is built for alignment to capture screening, diagnostics, and cost across an episode.

Acumen provided a recap of the discussions in the second TEP meeting in October 2023. During the October 2023 meeting, TEP members provided feedback on attribution, measure numerators and denominators, and time intervals between services for the quality measures. For the cost measure, the TEP discussed trigger codes, episode window, attribution, sub-populations of interest, and service assignment.

Acumen also recapped the input provided by the three PFPs on their experiences with breast cancer. The TEP was reminded of the PFPs' experiences with their screening mammography results, care team, types of services they received, indicators of high-quality care, and barriers to quality care. The PFPs' care teams included primary care physicians, surgeons, radiologists, nurse practitioners, oncologists, and insurance case managers. The types of services PFPs received included a screening mammogram, ultrasound, biopsy, chemotherapy, wound care, diagnostic mammogram, MRI, lumpectomy, radiation, and breast reduction surgery. The PFPs noted timeliness of results, good communication and coordination, and active shared decision-making as indicators of high-quality care, and cited inadequate guidance, access to care within siloed facilities, no shared decision-making, and scheduling and transportation as barriers to quality care.

2.2 Beta Testing Feedback Overview

Acumen conducted beta testing on the three quality measures and one episode-based cost measure from February 1 to February 26, 2024. Beta testing allowed for interested parties to

learn more about the draft measures and provide feedback on the draft specifications through an online survey. During beta testing, materials including draft measure specifications, preliminary measure testing results, and prior TEP meeting summaries were available on the CMS Measure Management System website. There was also a beta testing survey available that allowed stakeholders to provide feedback on the measures.

Acumen received responses from eight different individuals or organizations. Responses came from one professional medical society, two health systems, four teaching hospitals, and one breast clinic. Respondents provided specific feedback on measure specifications, which guided the discussion during the meeting and informed the TEP's decisions.

2.3 Quality Measures Discussion

The TEP revisited topics related to the quality measures specifications. Section 2.3.1 discusses a consideration to update to the terminology used to refer to "diagnostic follow-up." Section 2.3.2 discusses alignment between the Recall Rate measure and the Hospital OQR Program Breast Cancer Screening Recall Rates (OP-39) measure (hereafter referred to as Hospital OP-39 measure). Section 2.3.3 summarizes the discussion on follow-up windows for events downstream from a screening mammogram. Section 2.3.4 summarizes the discussion on whether or not to use E/M services with a breast cancer diagnosis billed by selected types of practitioners to indicate evidence of breast cancer. Lastly, Section 2.3.5 summarizes key takeaways from the above sections based on discussions during the meeting and poll results.

2.3.1 "Diagnostic Follow-Up" Terminology

Acumen received a comment during beta testing suggesting to use the term "recall from screening" or "diagnostic evaluation" instead of "diagnostic follow-up." The term "diagnostic follow-up" is currently used in the quality measures specifications to refer to additional imaging done after a screening mammogram, and include diagnostic digital breast tomosynthesis (DBT), diagnostic mammogram, ultrasound, and magnetic resonance imaging (MRI). The commenter noted that the term "diagnostic follow-up" might be misinterpreted as "looking at the finding again in 6 months-2 years instead of immediate action." The TEP members did not reach a consensus on whether the term "diagnostic follow-up" should be changed, so the current terminology of "diagnostic follow-up" in the quality measure specifications will be maintained.

2.3.2 Alignment Between the Recall Rate Measure and the Hospital OP-39 Measure

CMS considers alignment with related measures, so where possible, the clinician-level Recall Rate measure should consider alignment with the Hospital OP-39 measure's specifications. For reference, the Hospital OP-39 measure is a Breast Cancer Screening Recall Rates measure specified with claims-based proxies and used in CMS's Hospital OQR Program.

Acumen presented the similarities and differences between the Recall Rate measure and the Hospital OP-39 measure based on their measure specifications. Both the Recall Rate measure and the Hospital OP-39 measure (i) use a 45-day window between a screening mammogram and diagnostic follow-up and (ii) define a diagnostic follow-up as a diagnostic DBT, diagnostic mammogram, ultrasound of the breast, and MRI of the breast.

The measures differ in that the Recall Rate measure only includes women ages 40 years and older and excludes patients with a history of breast cancer on or 365 days prior to screening mammogram, while the Hospital OP-39 measure has no exclusions on age, sex, or conditions. The rationale behind these differences was decided in the previous TEP meetings and Acumen's review of coding guidance. The TEP previously voted to restrict the patient cohort of the Recall Rate to women 40 years and older to align with the U.S. Preventative Services Task Force guidelines that recommend screening women at the age of 40 years old. The TEP also previously voted to exclude patients with a history of breast cancer to ensure an average-risk patient screening population.

Acumen also presented information comparing the Current Procedural Terminology/Healthcare Common Procedure Coding System (CPT/HCPCS) codes used in both the Recall Rate measure and Hospital OP-39 measure and noted the differences between the codes used in each measure as well as the rationale for these differences. Both measures use the same codes to identify ultrasound of the breast, but the Hospital OP-39 measure uses additional codes for screening mammogram, diagnostic mammogram, diagnostic DBT, and MRI of the breast that our Recall Rate measure does not include. The TEP's considerations of including these additional codes in the Recall Rate measure are summarized below:

- **Screening mammogram codes:** The TEP agreed to not include codes that are specific to compliance and reporting of the measure.
- **Diagnostic mammogram codes:** The TEP agreed to not include codes that are not standalone codes, since they would be double counted if they were included in the Recall Rate measure. The TEP also did not want to include codes that are not screening tests.
- **Diagnostic DBT codes:** During the meeting, the TEP verbally agreed to include 3D breast mammography of both breasts in the Recall Rate measure for completeness, even if it is not currently paid for by Medicare. The TEP did not want to include single breast imaging codes since they were indicative of prior mammography and not the target population for the measure.
- **MRI of the breast codes:** The TEP previously agreed to not include codes defined as MRIs without contrast as they were not felt to be tests related to breast

cancer screening and diagnostics. During TEP meeting #3, there was verbal TEP consensus to include CPT/HCPCS codes for MRI services with contrast and MRI services both with and without contrast. When contrast was not specified in a code description, the TEP asked Acumen to check if it implies both with and without contrast, and if the code description includes both, then to add the code to the Recall Rate measure. The TEP was in favor of including codes where imaging was done for both breasts, and did not want to include codes of single breast imaging since it indicates prior screening.

2.3.3 Follow-Up Windows for Events Downstream from a Screening Mammogram

The TEP revisited the discussion on time intervals used in the definitions of the PPV1 and PPV3 measure numerators. Acumen discussed how shorter timeframes between the screening mammogram/diagnostic follow-up and breast cancer treatment encourage care coordination in a timely manner, and a longer timeframe allows for additional time for care coordination and treatment plans. Beta testing respondents noted that a longer time interval between diagnostic follow-up and breast cancer treatment recognizes delays that can occur due to patient concerns, staffing shortages, and resource or scheduling limitations. However, a longer window can cause patients emotional stress due to longer wait times for results. Acumen presented data on each of these timeframes to highlight the population-level pattern for the TEP to consider.

Time interval between screening mammogram and breast cancer treatment: For the PPV1 numerator, the current measure includes an 8-month time interval between a screening mammogram and breast cancer treatment. Some TEP members and PFPs expressed how important timeliness is for both the patient and the care provider and said eight months is sufficient. In some instances, they noted that eight months could be too long of a timeframe for patients with aggressive cancers. One TEP member noted that patients may leave practices or not return for follow-up if they have to wait too long. Other TEP members suggested extending this timeframe to a more inclusive time interval to capture patients that may take longer to receive treatment. One TEP member added that their older patients often delay treatment after diagnosis so they can take more tests to better understand their conditions. One of the TEP moderators suggested a longer time interval to make sure the measure is not miscounting screen-detected cancers that took longer to confirm as interval cancers. Beta testing feedback included seven votes for eight months and one vote for nine months. Based on the TEP discussions during the meeting and poll results, the TEP voted for an 8-month time interval between a screening mammogram and breast cancer treatment, as currently specified in the measure.

Time interval between diagnostic follow-up and breast cancer treatment: For the PPV3 numerator, the current measure includes a 4-month time interval between a diagnostic follow-up

and breast cancer treatment. One TEP member suggested a shorter time interval to reduce patient anxiety and stress. One of the TEP moderators suggested potentially a 3-month time interval based on the data showing that the median number of days between diagnostic follow up and breast cancer treatment is 39 days, and the average is 75 days. A shorter time interval captures the TEP's emphasis on coordinating a treatment plan as soon as possible for a diagnosed patient. Beta testing feedback included six votes for four months and two votes for five months. Since the TEP members did not reach consensus during the meeting or poll results, the current measure specification of a 4-month time interval between a diagnostic follow-up and breast cancer treatment will be maintained in the PPV3 numerator.

2.3.4 Types of Practitioners Billing E/M Services with a Breast Cancer Diagnosis

Currently, the PPV1 and PPV3 measure numerators capture breast cancer treatment services or at least two E/M services with a breast cancer diagnosis code as evidence of breast cancer. In the previous TEP meeting, TEP members suggested that there are inconsistencies in how primary care physicians might code a breast cancer diagnosis compared to specialists (such as an oncologist or radiologist).

Acumen recommended not limiting the types of practitioners that bill the E/M services with a breast cancer diagnosis included in the PPV1 and PPV3 numerators. Acumen presented an analysis showing that limiting the types of practitioners would result in a minimal difference in PPV1 and PPV3 rates because the majority of the evidence of breast cancer is provided by treatment-specific services, not by E/M services. Limiting the types of practitioners could also cause missed episodes and undercount screen-detected cancers that we may want to keep in the measure. However, not limiting the types of practitioners could result in overcounting. Based on beta testing feedback, five out of eight respondents voted against limiting the types of practitioners that bill E/M services with a breast cancer diagnosis. Some beta testing respondents recommended including E/M services if billed by radiologists, surgical oncologists, and radiation oncologists. One TEP member noted that by limiting the measure to practitioners billing E/M services with breast cancer treatment codes, the measure could miss patients that have a cancer diagnosis but do not receive treatment. TEP members generally agreed to maintain the current measures specifications and not limit the types of practitioners that bill E/M services with a breast cancer diagnosis.

2.3.5 Key Takeaways

Based on the TEP's discussion during the meeting and the poll results, the key takeaways for the quality measures include the following:

- The TEP did not reach a consensus whether or not to change the term "diagnostic follow-up," so the current terminology of "diagnostic follow-up" in the quality measure specifications will be maintained.

- The TEP agreed to not limit the types of practitioners billing E/M services with a breast cancer diagnosis.
- The TEP agreed to include additional CPT/HCPCS codes in the Recall Rate measure for 3D breast mammography of both breasts, and MRI services with contrast and MRI services both with and without contrast for both breasts.
- The TEP voted to retain the 8-month time interval between a screening mammogram and breast cancer treatment, as currently specified in the PPV1 numerator.
- The TEP did not reach consensus on the time interval between a diagnostic follow-up and breast cancer treatment in the PPV3 numerator, so the current measure specification of a 4-month time interval will be maintained.

2.4 Cost Measure Discussion

The TEP revisited topics related to the cost measure specifications. Section 2.4.1 discusses service assignment for the cost measure. Section 2.4.2 discusses options to incentivize timely cancer detection, either by subgrouping or removing costs from early detection episodes. Section 2.4.3 summarizes key takeaways from the above sections based on discussions during the meeting and poll results.

2.4.1 Service Assignment

The TEP revisited the service assignment rule for the cost measure, which aims to ensure that the measure incentivizes timely and early diagnosis by distinguishing between early and late breast cancer detection. For the purpose of the TEP discussion, Acumen referred to episodes with a breast cancer detection (based on the definition of evidence of cancer from the quality measures) within eight months of a screening mammogram as "early cancer detection" episodes, and those that are observed after eight months of the screening mammogram as "late cancer detection" episodes.

Currently, the cost measure assigns costs of breast cancer treatment only if breast cancer is detected after eight months of the screening mammogram. This time interval aligns with the time interval defined in the PPV1 measure numerator for alignment. Five out of eight respondents from beta testing agreed to assign treatment costs only if cancer is detected after the 8-month timeframe from a screening mammogram.

Additionally, Acumen presented feedback received during beta testing suggesting to assign costs of positron emission tomography/computed tomography (PET/CT) staging and cryoablation in the late detection services category after eight months of a screening mammogram. Some TEP members suggested that the costs of these services should not be assigned to the measure as these services are uncommon cancer services, PET/CT scans may be used in rare cases, and these services are not ordered by radiologists. However, there was also

discussion to include PET/CT scans in the measure since they may reflect geographic heterogeneity in practice and they may be used in certain cases, such as concern for internal mammary lymph nodes, triple-negative breast cancer, or stage four cancer with bone metastases. Acumen noted that it is helpful to understand all treatment practices so that the cost measure can highlight variation in care and also highlight cases when clinicians may be using non-guideline directed care.

2.4.2 Incentivizing Timely Cancer Detection

During the TEP meeting, Acumen presented two options to incentivize and reward early cancer detection: (i) creating subgroups based on breast cancer detection, or (ii) removing costs from the early cancer detection episodes.

Subgrouping based on breast cancer detection ensures that cancer episodes are compared to each other. Acumen presented a table showing that with subgrouping, early breast cancer detection episodes' risk-adjusted cost appears similar to no cancer episodes, and late cancer detection episodes' risk-adjusted cost still appear higher cost, but are less extreme. However, subgrouping by breast cancer detection may reduce performance gaps among providers by neutralizing the difference between no cancer episodes and early cancer detection episodes.

Assigning fewer services to the early detection episodes is another option to reward early detection. Acumen presented a table categorizing services as either "early detection or no detection" and "late detection" (Table 1). Based on the current service assignment rule, the costs of services in the "late detection" category would be assigned if breast cancer is detected after eight months of the screening mammogram. Clinicians with early detection or no detection episodes would only be assigned the costs in the basic diagnostic services category. Acumen noted that assigning fewer services to the early detection episodes may provide an incentive to successfully detect cancers (i.e., removing imaging cost). However, services not assigned may present an opportunity for over testing, and early detection cases may have a higher frequency of services than cases of no detection.

For the option of assigning fewer services to early detection episodes, some TEP members felt that it is unfair to only assign costs for late detection when early detection episodes could still require treatment services, and radiologists may be assigned treatments costs that are out of their control. Given some of these concerns, TEP members generally favored the option of subgrouping after Acumen discussed how subgrouping could distinguish between radiologists who are detecting cancer early and those who are finding interval cancers, addressing previous concerns that radiologists were not receiving sufficient credit for early detection and that the measure was overattributing costs of cancer care to the radiologists.

Table 1. Service Assignment Rule for No, Early, and Late Breast Cancer Detection

EARLY DETECTION OR NO DETECTION		
Basic Diagnostic Services	• Mammography • Diagnostic ultrasound • Breast biopsy	• MRI • E/M (encounter for screening mammogram)
LATE DETECTION		
Basic Diagnostic Services	• Mammography • Diagnostic ultrasound • Breast biopsy	• MRI • E/M (encounter for screening mammogram)
Advanced Diagnostic Services	• Pathology (including surgical pathology)	• Genetic testing
Treatment Services	• Chemotherapy • Lumpectomy, quadrantectomy of breast • Mastectomy • Therapeutic radiology • Anesthesia • Laboratory – Chemistry and Hematology • Non-hospital-based care	• Ancillary services • Medications (injections, infusions, and other forms) • DME and supplies • Hospitalizations (malignant breast disorders, septicemia or severe sepsis) • Complications of treatment (including hemorrhage) • E/M (breast cancer diagnosis)

2.4.3 Key Takeaways

Based on the TEP's discussion during the meeting and the poll results, the key takeaways for the cost measure include the following:

- The TEP favored subgrouping by breast cancer detection as a method of rewarding early detection.
- The TEP agreed to assign PET/CT staging costs to the late detection services category.
- The TEP had diverging views on whether cryoablation costs should be assigned to the late detection services category, and ultimately did not reach a consensus. Upon reviewing the beta testing comments, reviewing coding manuals and guidelines, considering the viewpoints expressed during the meeting, and consulting with the Acumen clinical team, Acumen will assign cryoablation costs to the late detection services category, as cryoablation is in fact used for breast cancer treatment, albeit infrequently.

2.5 Cross-Measure Discussion

The TEP had a cross-measure discussion that involved both the quality and cost measures. Section 2.5.1 discusses the proposed measure scoring methodology. Section 2.5.2 covers the key takeaways from the discussion on the measure scoring methodology during the meeting and the poll results. Section 2.5.3 discusses the findings from an analysis on social risk factors.

2.5.1 Measure Scoring Methodology

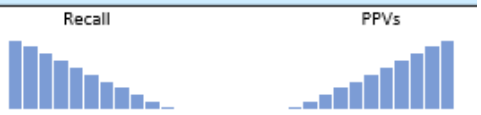
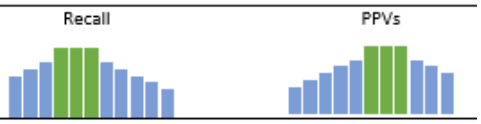
In the MIPS program, providers are scored based on their ranking relative to their peers. Currently, the quality measure scores reflect the percentage rates. It is generally agreed upon that there is societal consensus on ranges for these metrics, but higher or lower rates do not necessarily indicate higher quality care. Two beta testing respondents agreed that it is important to consider the acceptable ranges. Other beta testing feedback noted considerations in benchmarking: variation in screening patterns of 40- to 50-year-old women is not accounted for; highly specialized centers may have higher risk populations and higher call backs and needs for biopsies; and rates are also influenced by factors, such as access to care, frequency of screenings, quality of imaging, clinician experience, age, and screening intervals.

To prepare for use in MIPS, Acumen proposed four benchmarking methodologies to aid in the ranking of these measure scores. These methodologies consider three dimensions for adjusting performance ranking:

- Performance within the acceptable range is better
- Differentiation within the acceptable range, by assuming lower Recall is better and higher PPVs are better
- Symmetry of penalization outside of the acceptable range (i.e., +1 and -1 from acceptable range are similar), or penalize higher than acceptable Recall more and lower than acceptable PPVs more

Acumen presented a figure (Figure 1) showing a simulated range of scenarios for the quality measures based on these adjustments, and asked the TEP which combination would be most appropriate for scoring. The green bars indicate scores within the acceptable range, and the blue bars indicate scores falling outside of the acceptable range.

Figure 1. Proposed Benchmarking Methodologies

Option	(i) Performance within the acceptable range is better	(ii) Differentiation within the acceptable range	(iii) Symmetry of penalization outside of the acceptable range	Ranking Illustration acceptable range
Baseline	No	No	No	
A	Yes	No	Symmetrical	
B	Yes	Yes, lower Recall is better and higher PPVs are better	Symmetrical	
C	Yes	Yes, lower Recall is better and higher PPVs are better	Penalize higher than acceptable Recall more and lower than acceptable PPVs more	
D	Yes	No	Penalize higher than acceptable Recall more and lower than acceptable PPVs more	

Acumen also showed the TEP a simulation of how providers with various performance characteristics may perform across the Recall, PPV1, and PPV3 measures with each performance ranking adjustment applied. The final part of the simulation showed the rankings with the quality measures and cost measure combined. The simulation showed:

- Ranking using raw rates (i.e., baseline option) does not reward providers who are within the acceptable range on all three quality measures
- Providers within the acceptable range always rank highest across options A-D to reward balanced performance
- Options A and D may lead to a topped-out measure where many providers get the maximum score due to not having differentiating within the acceptable range, reducing the ability to differentiate across providers
- Further differentiation between providers outside the acceptable range is observed in options C and D
- All methods consistently reward cancer detection
- When looking at the mean combined ranking of both the quality and cost measures (i.e., when the cost measure is included), late cancer detection is penalized

The TEP discussed the different methodologies. Several TEP members agreed that the PPV1/PPV3 rates and Recall rate are not of equal concern for clinicians in practice, and that missing cancer detection is a bigger concern than a higher than acceptable Recall rate. One TEP member said that the PPV rate is important, and that a higher Recall rate with a low cancer detection rate is preferable to a lower Recall rate with a lower cancer detection rate to ensure that no cancers are missed. Another TEP member said that for clinicians within the acceptable range, lower Recall is not always better since it can sometimes result in missed cancers because it could be associated with lower cancer detection and lower PPVs.

One TEP member said that for clinicians outside of the acceptable range, higher cancer detection rate and higher PPVs are much more preferable than a lower Recall rate at the expense of missing cancers. This TEP member suggested weighing the Recall Rate and PPVs differently by using the Recall graph from Option A and the PPV graph from Option C, since Option C penalizes scores outside the acceptable range to encourage clinicians to catch more cancers. This TEP member favored Option A for Recall over Option D because it does not penalize radiologists as much for calling back more cancers compared to not recalling enough cancers. Several TEP members agreed with this combination of methodologies.

Acumen noted that there is a need to differentiate within the range for Recall. Acumen explained that a challenge of not differentiating within the range for Recall (Options A or D) is that currently, 58% of providers are within the American College of Radiology (ACR) benchmark range, meaning that 58% of providers would receive a score of 100. This could result in a topped-out measure, making it more difficult to get the measure implemented into a CMS program. Acumen emphasized the importance of feasibility in getting this measure accepted into MIPS, and asked the TEP to consider Option B for the Recall measure. One of the moderators added that it is harder to not recall than to recall, and the risk of the topped-out measure is losing the ability to help close the gap in care of 42% of clinicians being outside of the range.

2.5.2 Key Takeaways

Based on the TEP's discussion during the meeting and the poll results, the key takeaways for the measure scoring methodology include the following:

- The TEP ultimately did not reach consensus on which options to select for each quality measure. Upon considering the (i) viewpoints expressed during the meeting, (ii) data presented during the meeting showing providers' performance rankings based on each proposed scoring methodology option, (iii) feasibility of the measures' potential implementation into a CMS program, and (iv) comments received during beta testing, Acumen will recommend to CMS Option B for the Recall Rate measure, Option C for the PPV1 measure, and Option C for the PPV3 measure.

2.5.3 Social Risk Factor Analysis

Beyond clinical characteristics of patients, the quality of care may be influenced by non-clinical factors related to a patient's social risk factors (SRFs), such as race or socioeconomic status. At the program level, MIPS adjusts for SRFs using the MIPS Complex Patient Bonus to ensure clinicians or groups treating more complex patients are not disadvantaged. Dual Medicare and Medicaid enrollment status was the focus of Acumen's analysis, as the data is available for all beneficiaries and dual enrollment status tends to be the most powerful predictor of poor outcomes.

The data analysis showed that when looking at performance stratified by clinician's dual share and patient's dual status for the quality measures, there were small differences between dual episodes and non-dual episodes, and the small variation in quality measure rates does not translate to a systemic difference in performance at the individual provider level.

When looking at performance stratified by clinician's dual share and patient's dual status for the cost measure, there is negligible variation between dual episodes and non-dual episodes at the both the group and individual reporting level. The data showed that most clinicians perform similarly on their dual and non-dual patients for both the quality and cost measures. Acumen explained that based on the results of the data analysis, it is not recommended to risk adjust for dual status in the quality and cost measures.

2.6 Preliminary Face Validity Vote

Acumen conducted a live poll during the TEP meeting to serve as a preliminary vote for face validity, which will help Acumen identify areas to refine the measures before they are finalized. Acumen may follow up with another face validity vote for further feedback. Acumen asked the TEP eight questions, with four on whether each measure can capture performance on quality or resource use if the specifications were updated based on the TEP's discussion, and four on if each measure can distinguish good from poor performance with the updated specifications based on the TEP's discussion and the suggested measure scoring methodology. The TEP chose from answers ranging from strongly agree, agree, undecided, disagree, and strongly disagree.

Table 2 shows the poll results for the first four questions on whether each measure can capture performance on quality or resource use if the specifications were updated based on the TEP's discussion. Table 3 shows the poll results for the other four questions on if each measure can distinguish good from poor performance with the updated specifications based on the TEP's discussion and the suggested measure scoring methodology.

Table 2. Preliminary Face Validity Results Based on Updated Measure Specifications

Measures	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
Recall Rate	37.5%	62.5%	0%	0%	0%
PPV1	37.5%	62.5%	0%	0%	0%
PPV3	25%	75%	0%	0%	0%
Cost	12.5%	75%	12.5%	0%	0%

Table 3. Preliminary Face Validity Results Based on Updated Measure Specifications and Measure Scoring Methodology

Measures	Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree
Recall Rate	12.5%	75%	12.5%	0%	0%
PPV1	0%	100%	0%	0%	0%
PPV3	12.5%	87.5%	0%	0%	0%
Cost	12.5%	75%	12.5%	0%	0%

3 NEXT STEPS

After the meeting, TEP members will receive a link to the recording as well as a poll to vote on measure specifications. Acumen will use the TEP's input from the meeting discussions and poll results to refine and finalize the draft measures. Acumen intends to submit the measures for the Measures Under Consideration (MUC) List consideration in May 2024.

If you have any questions, please contact the Acumen Mammography Measures Moore Support Team at mmg_measures_moore@acumenllc.com.

APPENDIX A: LIST OF TEP MEMBERS AND ACUMEN PROJECT TEAM

The *Improving the Diagnostic Performance of Screening Tests for Breast Cancer* Technical Expert Panel is made up of 14 members (11 clinical, three person and family partners). Of the 14 members, 11 were able to attend the March 8, 2024 meeting. Section A.1 lists the clinical members. Section A.2 lists the person and family partner members. Section A.3 lists the 12 Acumen project team members for additional reference.

A.1 TEP Members: Clinical

- Megan Adamson, MD, MHS-CL, FAAFP, DipACLM, DipABOM
- Jose Gilberto Bazan, MD, MS
- Stamatia Destounis, MD, FACR, FSBI, FAIUM
- Carolyn Dueñas, RN, MBA, *Absent*
- Sarah Eakin, MD, FCAP
- Sharad Goyal, MD, MS, *Absent*
- Cindy Lee, MD, FSBI, CMQ
- Lauren Nicola, MD, *Absent*
- Lydia Pace, MD, MPH
- Barbara Spivak, MD
- Barbara Wexelman, MD, MBA

A.2 TEP Members: PFPs

- Rosie Bartel
- Nancy McMahan Farrar
- Barbara Kivowitz

A.3 Acumen Project Team

- Rose Do, MD, Co-Principal Investigator
- David Seidenwurm, MD, FACR, Co-Principal Investigator
- Heather Litvinoff, PT, MPH, Project Manager *(Moderator)*
- Sri Nagavarapu, PhD, Technical Analytic Advisor
- Lois Olinger, MCP, Senior Policy Advisor
- Laurie Feinberg, MD, MPH, Clinical Associate
- Joyce Lam, MPP, Research Manager
- Ken Tran, PhD, Senior Policy Associate
- Kevin Lei, MS, Senior Policy Researcher
- Sarah Sabbagh, MPH, Policy Associate
- Suzie Choi, BS, Data & Policy Analyst
- Alexis Shannon, BA, Administrative Assistant
- Julia Lotan, BA, Data & Policy Analyst