Partnership for Quality Measurement
Powered by Battelle

National Consensus Development and Strategic Planning for Health Care Quality Measurement

Fall 2024 Cycle Endorsement and Maintenance (E&M) Meeting Discussion Guide

INITIAL RECOGNITION AND MANAGEMENT COMMITTEE

February 12, 2025

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

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.

Table of Contents

List of Tables

List of Figures

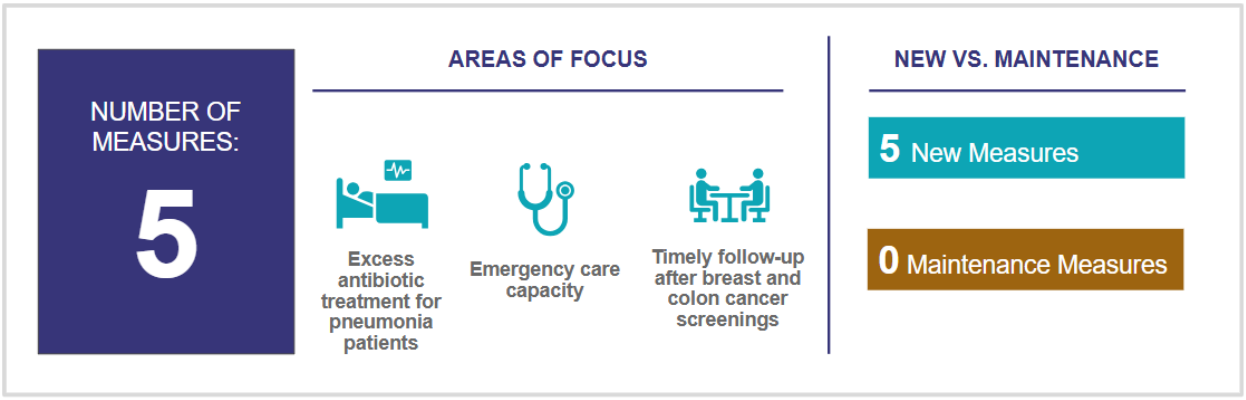
Overview of Fall 2024 Measures for Review

During this measure review cycle, developers and stewards submitted six measures to the Initial Recognition and Management committee and one measure was withdrawn by the developer (CBE #4720), leaving five measures for endorsement consideration (Table 1). The measures focused on excess antibiotic treatment for pneumonia patients, emergency care capacity, and timely follow-up after breast and colon cancer screenings (Figure 1).

Table 1. Overview of Measures Under Endorsement Review

CBE Number	Measure Title	New/Maintenance	Developer/Steward
4540e	Excess Antibiotic Duration for Adult Hospitalized Patients with Uncomplicated Community-Acquired Pneumonia	New	University of Utah
4545e	Inappropriately Broad Empiric Antibiotic Selection for Adult Hospitalized Patients with Uncomplicated Community-Acquired Pneumonia	New	University of Utah
4625e	Emergency Care Capacity and Quality eCQM	New	Acumen/Centers for Medicare and Medicaid Services (CMS)
4700e	Rate of Timely Follow-up on Abnormal Screening Mammograms for Breast Cancer Detection	New	Brigham and Women's Hospital
4705e	Rate of Timely Follow-up on Positive Stool-based Screening Tests for Colorectal Cancer Detection	New	Brigham and Women's Hospital

Figure 1. Fall 2024 Measures for Committee Review

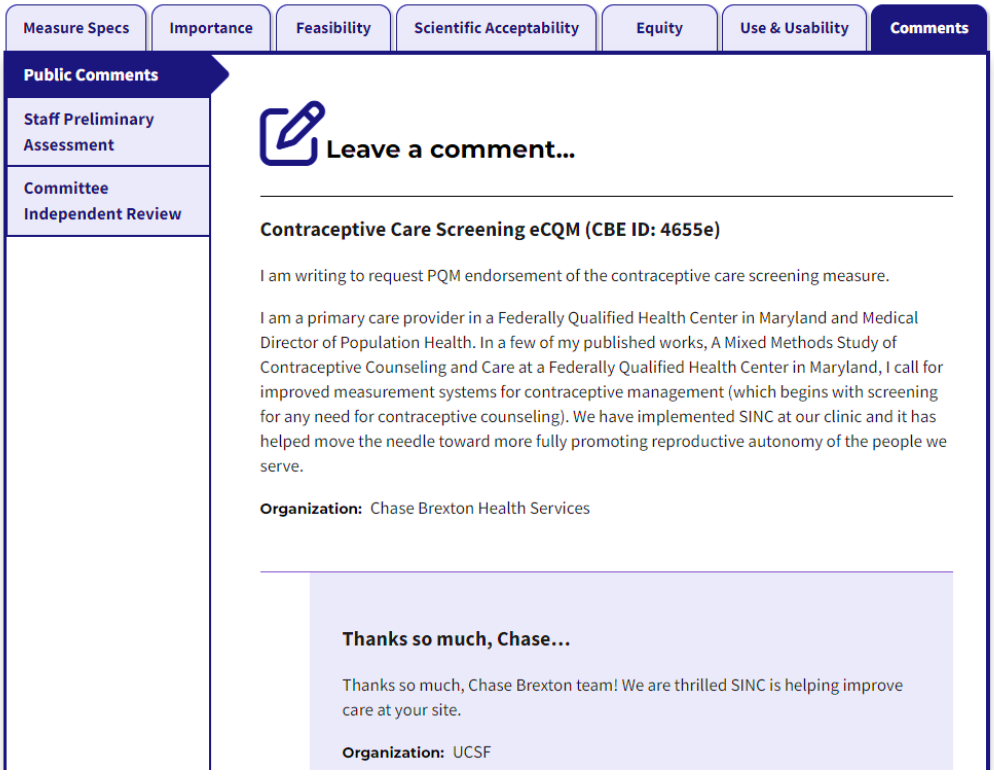


Public Comment

Battelle accepts comments on measures under endorsement review through the Partnership for Quality Measurement (PQM) website and Public Comment Listening Sessions. In this evaluation cycle, the public comment period opened on November 15, 2024, and closed on December 16, 2024. Battelle held a Public Comment Listening Session on November 21, 2024.

After the public comment period closed, developers/stewards had the opportunity to respond to public comments on the measure page in the Submission Tool and Repository Measure Database (STAR). To view the public comments and response, go to the "Comments" tab in the left navigation pane (Figure 2). Each comment has a bold heading followed by the body of the comment. Developer responses, if any, appear as a shaded reply beneath the comments. Note that developers are not obligated to respond to public comments. Lastly, the measure evaluation summaries below contain the number of public comments received for each measure.

Figure 2. Viewing Public Comments and Developer Responses



Advisory Group Feedback

The Advisory Group convened on December 5, 2024; 24 of 30 (80%) active Advisory Group members attended to share feedback and ask questions regarding the measures under endorsement review. Developers/stewards of the respective measures also attended and provided responses to the Advisory Group questions. After the meeting, developers/stewards had the opportunity to submit additional written responses to Advisory Group member feedback and questions (Appendix A).

The measure evaluation summaries of this discussion guide contain overviews of the Advisory Group member discussions and developer/steward responses.

To support the review of the public comments and Advisory Group summaries, the number of comments received or number of individuals who shared similar comments, feedback, and/or questions is represented as “a few” (two to three individuals), “several” (four to six individuals), and “many” (more than six individuals). This discussion guide also employs four key categories—Supportive, Dissenting, Mixed, and Probing—to structure and enhance the Recommendation Group discussion.

- **Supportive**: This includes views and comments that express agreement, encouragement, or reinforcement of the measure.
- **Dissenting**: This captures opinions that disagree with or oppose what has been stated about the measure or what has been provided within the measure submission.
- **Mixed**: This category encompasses feedback that contains both supportive and dissenting elements.
- **Probing**: This involves questions or comments that seek to explore, clarify, or delve deeper into aspects of the measure.

Measures Under Endorsement Review

CBE 4540e: Excess Antibiotic Duration for Adult Hospitalized Patients with Uncomplicated Community-Acquired Pneumonia [University of Utah]

Specifications

Measure Description: *The Excess Antibiotic Duration for Adult Hospitalized Patients with Uncomplicated Community-Acquired Pneumonia measure is a process measure representing the annual percentage of hospitalized adults with uncomplicated community-acquired pneumonia who receive an excess antibiotic duration. The measure will be calculated using electronic health record (EHR) data and is intended for use at the facility level for both quality improvement and pay-for-performance.*

Staff Preliminary Assessment Rating[1]

Importance: Not Met but Addressable

Rationale: This new measure is supported by a clear logic model linking inputs like ATS/IDSA Guidelines and EHR Systems to activities and outcomes such as improved guideline adherence and patient safety, addresses a gap in existing measures for CAP, as evidenced by variable performance data and a comprehensive literature review. However, the information about meaningfulness to patients is limited to evidence from guidelines and a TEP that included only clinicians.

Feasibility: Met

Rationale: This new measure meets all criteria for “Met” due to its well-documented feasibility assessment, clear and implementable data collection strategy, ensuring practical implementation within the healthcare system.

The eCQM Feasibility Scorecard results indicate that the measure is well-supported by the current capabilities of the tested EHR systems in terms of data availability, accuracy, adherence to standards, and workflow integration. These factors collectively ensure that the measure can be implemented effectively and sustainably in a real-world healthcare setting.

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the accountable entity level.

Validity: Not Met but Addressable

Rationale: The validity testing results support a relatively strong inference of validity for the measure, confirming that the measure accurately reflects performance on quality or resource use and can distinguish good from poor performance. However, additional data element testing within at least two EHR vendors is needed.

Equity: Not Met but Addressable

Rationale: Though the developer provided some information on inequities, no empirical testing was completed.

[1] Located under the “Comments” tab, then “Staff Preliminary Assessment.”

Use & Usability: Met

Rationale: This new measure is currently used in at least one accountability application, and the measure provides actionable information for improvement. The developer describes approaches to collecting feedback on this measure, but more detail is needed on how that feedback led to changes in the measure specifications, especially for maintenance endorsement review. Although performance results for the chart-based version of the measure showed positive results, more data on performance trends for the eCQM is needed. The developer reports no unexpected findings.

Public Comment

Number of Comments Received During the Public Comment Period: 7

Comments and their responses from measure developers can be found on the measure page under the "Comments" tab (*Figure 2*).

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Excess Antibiotic Duration Definition: A few patient participants of the Advisory Group asked for clarification on how "excess antibiotic duration" is defined, stating that this wasn't immediately clear to them. An Advisory Group member asked if "duration" is considering consecutive days or total days per visit.	The assessment is based on how long it takes a patient to become clinically stable. Based on clinical trials, anything beyond 5 days is considered excess; however, 7 days is used to allow for potentially missed coding or a half day off for antibiotic duration. Duration is the number of days in which a patient receives antibiotics during hospitalization plus the prescribed discharge antibiotic duration.
Uncomplicated Pneumonia Definition: An Advisory Group member asked for clarification on how "uncomplicated pneumonia" is defined and how often patients with uncomplicated pneumonia are admitted.	The developer is trying to home in on patients who do not require intensive care unit (ICU)-level care, have a severe comorbidity, or develop unusual pathogens or pathogens that require longer treatment. While they do not know the percentage of patients with uncomplicated pneumonia who are admitted, they do know that in patients who are hospitalized, pneumonia is the number one predictor of receiving antibiotics and count for 40% of antibiotics prescribed during hospitalization. They noted that this did vary by hospital, with smaller hospitals being more prone to overuse.
Exclusion Timeframes: An Advisory Group member asked for more information on the exclusions related to if a patient is on antibiotics for less than 3 days or greater than 14 days.	The developer decided on the 3-day and 14-day exclusions based on chart review. Patients who are treated for pneumonia with antibiotics for less than 3 days usually have stopped that course of treatment because it has been found that they do not actually have pneumonia. Patients who are being treated for pneumonia with antibiotics for greater than 14 days tend to have had a clinical reason or complication for having been on antibiotics for such an extended duration of time.

Feedback/Questions	Summary of Developer Response
Inpatient versus Observation: An Advisory Group member asked if the measure includes both inpatient and observation individuals.	The measure includes both because, in part, what those patients are called varies across hospital. The developer did not want variation in coding to affect quality of care.
Inpatient versus Outpatient Prescriptions: An Advisory Group member asked if the measure accounts for outpatient prescriptions for antibiotics in addition to inpatient prescriptions?	Both discharge antibiotics and antibiotics given during hospitalization are included. Most excess duration happens after discharge.
Coding Intensity: An Advisory Group member commented that with such a variety of places, they would expect see different coding intensity.	The developer discussed at length how to identify patients at admissions. They said they realized that many facilities were practicing upcoding and bill for sepsis and respiratory failure. To accommodate for that, the developer inclusion criteria include individuals coded with pneumonia or individuals coded for sepsis and respiratory failure.
Sepsis and Respiratory Failure: A few Advisory Group members stated that they found the wording of the denominator confusing, believing that the measure would include pneumonia, sepsis, or respiratory failure. A few Advisory Group members also stated that while they understood the developer's intention to catch instances of pneumonia where a facility might be "upcoding," (i.e., using a billing code that reflects a more severe diagnosis or more extensive procedure than what was actually provided to the patient), they struggled with sepsis and respiratory failure being considered uncomplicated.	The developer did not specifically respond to the comment about wording. Regarding sepsis and respiratory failure, the issue is the upcoding, and that they did not want those patients to be missed simply because a facility is upcoding.
Coding Discrepancies: An Advisory Group member pointed out that coding rules do not always align with clinical practice when it comes to complicated versus uncomplicated pneumonia. They asked if the developer could rely on anything in addition to ICD-10 codes. A few Advisory Group members commented that this measure may encourage facilities to be more careful and accurate with their coding information. One Advisory Group member said that it might be helpful to provide education related to the coding for this measure.	In part, this is why they excluded patients who were treated for more than 14 days. They added that they anticipate that facilities will need to code the correct items, as is the case with many measures. They highlighted that, when compared to chart review, the measure has sensitivity and specificity of over 95% for identifying accurate assessment of excess duration.
Validity: An Advisory Group member said that within the submission material's validation studies, it appeared that 30-50% fell outside of the range of acceptable, which seemed high to them.	The measure to has 96% sensitivity and 93% specificity in looking at appropriate duration, meaning that they were fairly accurate.
Exclusions: A few Advisory Group members commented that the exclusions looked reasonable to them and would result in the measure capturing what it intended.	N/A
Drug Allergies: An Advisory Group member said they did not believe multiple drug allergies would be a confounder for this measure.	Allergies are not applicable to this measure, although they are for #4545e.

Feedback/Questions	Summary of Developer Response
Value Sets: An Advisory Group member expressed concern that the value sets were not achieving what the developer believed. Specifically, they highlighted: • The denominator has a single value set; however, given that measure is capturing any individual who is admitted for community-acquired pneumonia or who is admitted for sepsis and respiratory failure, they expected to see three different value sets. • The antibiotic list does not appear to include macrolides and tetracyclines. However, Zithromax doxycycline is a common treatment for uncomplicated pneumonia. If these were intentionally not included, the Advisory Group member would like to know why. • The comorbidities value set does not do what the developer describes. • Mechanical ventilation is not represented in the value sets. The Advisory Group member also encouraged the developer to align with the IDSA if possible.	For the sepsis value set and the respiratory value set comment, the value sets were accurate but were combined when they should have been separated. The developer has now divided this into three value sets: CAP Pneumonia Diagnostic, CAP Sepsis Diagnostic, and CAP Respiratory Failure diagnostic. To enter into the measure denominator, a patient must have either: a) a pneumonia code (by itself) or b) a sepsis code AND a respiratory failure code. For the antibiotics list, the developer apologized for this oversight. Macrolides and doxycycline are not included in "AntibioticUsageforCAP" purposely because patients treated with azithromycin alone or doxycycline alone commonly have a chronic obstructive pulmonary (COPD) exacerbation, and the developer did not want those patients included in the denominator. However, the developer acknowledged that they meant to include those antibiotics in counting duration for antibiotics. They have created a second value set "AntibioticUsageforCAPduration" that includes these antibiotics (and others) to ensure they are accurately counting duration. For the comorbidities value set, the developer realized they had kept in an outdated set. The ComorbiditiesIndicatedwithCAP value set has now been updated to match the specifications listed in the measure submission. No coding changes required. Please see the updated data dictionary for details. For the mechanical ventilation, the CPT codes were not accurate at identifying mechanical ventilation. Thus, they instead excluded patients admitted to the ICU (where ventilated patients are normally managed), excluded patients with complications of ventilation, and excluded patients with a tracheostomy. The tracheostomy and mechanical ventilation complications are found in the comorbidities value set. The MajorTransplant value set lists a comprehensive list of solid organ transplant codes to identify immunosuppressed patients and has been revised (with the value set: "TransplantsStemCellandSolid") to also include codes associated with stem cell transplants.

Feedback/Questions	Summary of Developer Response
Present-on-Admissions Information: An Advisory Group member commented that it was not clear how the developer was incorporating present-on-admission information to ensure that they are focusing on community-acquired pneumonia rather than hospital-acquired pneumonia.	Antibiotics are required to be administered on day 1 or day 2 to ensure they are dealing with community-acquired pneumonia rather than hospital-acquired pneumonia. The developer's rationale is that if the patient is receiving antibiotics that soon, they are being treated for something that was present on admission.
Balancing Accuracy vs Feasibility: An Advisory Group member commented that the developer addressed 29% sensitivity on whether the patients have improved enough within 5 days to go off antibiotics by simplifying some of the criteria, including fever, hypotension criteria, and hypoxia criteria. The Advisory Group member asked if the developer felt okay with this decision.	The developer spent a lot of time considering how to balance accuracy and feasibility. They found that using the modified definition in their University of Michigan population yielded similar results (with the expanded vital sign definition, 53.7% were considered excess duration whereas with the modified definition, 54.5% were considered excess duration). They talked to their TEP, and the TEP felt the full version would require additional work from the facilities. They said the tradeoff of a small amount of accuracy for improved feasibility seemed appropriate.
Immunocompromised Exclusions: An Advisory Group member pointed out that the Infectious Disease Society of America's guidelines are broader for immunocompromised patients who do not qualify for the 5-day criteria. They asked how the developer reached their exclusion criteria for immunocompromised patients.	The developer could not find clear guidance in the guidelines, so they went to the original study in JAMA Internal Medicine by Uranga and prioritized their definitions. The developer found extreme variation in how facilities coded mild and moderate immunocompromised patients, so they focused on excluding the most severe immunocompromised patients.
Procalcitonin: An Advisory Group member commented that procalcitonin can be used to identify patients with severe pneumonia; however, they added that procalcitonin can be negative or normal in patients with atypical bacteria pneumonia and may not be appropriate to for excluding patients from the measure.	The developer did not take procalcitonin into effect because of variation in practice and mixed guidance on whether or not it should be used in the United States.
Positive Cultures: An Advisory Group member asked how the measure accounts for that pneumonia does not typically have positive cultures.	While most pneumonia is culture negative, trials also indicate that if patients are culture negative, do not improve, and do not quality for complicated pneumonia, they should receive a 5-day antibiotic treatment. They highlighted that they are advocating for patients to be treated based on empirical evidence.
Severity of Infection: An Advisory Group member commented on how severity does not seem be parsed out in this measure, noting that the IDSA has guidelines.	This measure is based on time to stability, and that there is less evidence that severity is noteworthy for a duration-based measure. If a disease is severe enough, an individual would go to the ICU, which is one of the exclusion criteria for this measure. They added that they use a modified version of the severity criteria for #4545e.
Feasibility: An Advisory Group member said, given the list of exclusions, the measure will likely be burdensome to hospitals that will need to produce a parallel measure to track this information.	Making the measure an eCQM will make it more useful to a broader swathe of hospitals. They also stated that they have been talking with Epic about putting this on their dashboard.

Feedback/Questions	Summary of Developer Response
Evidence: An Advisory Group member praised the measure, stating the antibiotic treatment used to be decided based on "gut feeling" and that the length of treatment the measure suggests is supported by a large amount of evidence. They added that longer courses of antibiotics increase cost, time in hospital, and antibiotic resistance.	N/A
Smaller Facilities: Committee members discussed how this measure may impact smaller facilities. One Advisory Group member stated that this measure may be beneficial to them. Another stated that other smaller hospitals may admit more individuals to the ICU to "get around" the measure.	The developer did not specifically address this comment.
Proposed versus Required Measure: An Advisory Group member asked, as an eCQM, is this proposed as an elective option, or is the intent for this measure to become a required measure?	This measure be initially a pay-for-reporting measure, similar to the National Healthcare Safety Network Antimicrobial Use and Resistance (NHSN AUR) measure. Just having the data will enable stewardship teams to direct improvement. Battelle Staff Note: Pre-Rulemaking Measure Review (PRMR) is a process separate from endorsement and maintenance and is used to make recommendations on measure use within a CMS quality reporting or value-based program.

Key Themes from Advisory Group Feedback, Public Comment, and Staff Assessments

Discussion Categories	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance and Evidence	Advisory Group; Public Comment	An Advisory Group member praised the measure for its basis in strong evidence, stating that the measure could help reduce cost, length of hospitalization, and antibiotic resistance. Public comments received on this measure further support its importance, expressing that the measure is interoperable and aligns with recommendations from organizations like the CDC, The Joint Commission, and CMS.
	Exclusions	Advisory Group	A few Advisory Group members stated that the exclusion criteria seemed sound for trying to identify the appropriate patient population.
Dissenting	Value Sets	Advisory Group	An Advisory Group member stated that several of the value sets were not functioning in the matter in which the developer described.
	Feasibility	Advisory Group	Despite this being an eCQM, an Advisory Group member stated that they believed this will still be a burdensome measure for facilities due to the lengthy exclusions list. A few Advisory Group members also

Discussion Categories	Key Themes	Source of Comment	Summary of Comments
			expressed concern over how accurately coding would reflect clinical practice.
	Validity Testing	Staff Assessment	The validity testing results support a relatively strong inference of validity for the measure, confirming that the measure accurately reflects performance on quality or resource use and can distinguish good from poor performance. However, additional data element testing within at least two EHR vendors is needed.
Mixed	Denominator	Advisory Group	Advisory Group members discussed the inclusion of sepsis and respiratory failure in the measure's denominator. They indicated that while they understood the developer's intention to catch instances of pneumonia where a facility might be "upcoding," it still felt odd to see sepsis and respiratory failure as being considered "uncomplicated." They also added that, as how the denominator is currently worded, they misunderstood that the measure includes 1) pneumonia or 2) sepsis and respiratory failure, believing that individuals could be included for 1) pneumonia, 2) sepsis, or 3) respiratory failure.
	Meaningfulness to Patients	Staff Assessment; Public Comment	Meaningfulness to patients was partially assessed. The information available in the literature was limited and the TEP only included clinicians. However, a public comment shared by Patients for Patient Safety US, an advocacy group, expressed the importance of CBE #4545e and #4540e.
Probing	Equity	Staff Assessment	Though the developer provided some information on inequities, no empirical testing was completed.

CBE 4545e: Inappropriately Broad Empiric Antibiotic Selection for Adult Hospitalized Patients with Uncomplicated Community-Acquired Pneumonia [University of Utah]

Specifications

Measure Description: *The Inappropriately Broad Empiric Antibiotic Selection for Adult Hospitalized Patients with Uncomplicated Pneumonia measure is a process measure representing the annual percentage of hospitalized adults with uncomplicated community-acquired pneumonia. Here, we defined "inappropriately broad" as any antibiotic therapy targeting methicillin-resistant Staphylococcus aureus (MRSA) or Pseudomonas aeruginosa in patients without risk factors for one of those organisms. The measure will be calculated using electronic health record (EHR) data and is intended for use at the facility level for both quality improvement and pay-for-performance.*

Staff Preliminary Assessment Rating

Importance: Not Met but Addressable

Rationale: This new measure is supported by a clear logic model linking inputs like ATS/IDSA Guidelines and EHR Systems to activities and outcomes such as improved guideline adherence and patient safety, addresses a gap in existing measures for CAP, as evidenced by variable performance data and a comprehensive literature review. However, the information about meaningfulness to patients is limited with evidence from guidelines and a TEP that included only clinicians.

Feasibility: Met

Rationale: This new measure meets all criteria for "Met" due to its well-documented feasibility assessment, clear and implementable data collection strategy, ensuring practical implementation within the healthcare system.

The eCQM Feasibility Scorecard results indicate that the measure is well-supported by the current capabilities of the tested EHR systems in terms of data availability, accuracy, adherence to standards, and workflow integration. These factors collectively ensure that the measure can be implemented effectively and sustainably in a real-world healthcare setting.

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the accountable entity level.

Validity: Not Met but Addressable

Rationale: The validity testing results support a relatively strong inference of validity for the measure, confirming that the measure accurately reflects performance on quality or resource use and can distinguish good from poor performance. However, additional data element testing within at least two EHR vendors is needed.

Equity: Not Met but Addressable

Rationale: Though the developer provided some information on inequities, no empirical testing was completed.

Use & Usability: Met

Rationale: This new measure is currently used in at least one accountability application, and the measure provides actionable information for improvement. The developer describes approaches to collecting feedback on this measure, but more detail is needed on how specific feedback received led to changes in the measure specifications. Although performance results for the chart-based version of the measure showed positive results, more data on performance trends for the eCQM is needed. The developer reports no unexpected findings.

Public Comment

Number of Comments Received During the Public Comment Period: 9

Comments and their responses from measure developers can be found on the measure page under the "Comments" tab (*Figure 2*).

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Importance: Several Advisory Group members highlighted that they believed this was an important measure, with a few adding that they wanted to emphasize to the committee that it was important to not throw away the benefits of the measure in search of absolute perfection. They expressed the belief that this measure would be valuable in educating providers.	This measure is important for reducing patient risk of antibiotic resistance. The developer has seen impact in the systems in Michigan that have used the measure, and that the measure is highly sensitive and specific when compared to chart review for identifying prior hospitalization and prior isolation of specific cultures. With how common overuse is, there is a lot of room for improvement.
Accessibility: A few patient participants on the committee stated that the measure's language was complicated and was not easily understood by laypeople. They asked for more information about the measure as well as an explanation for how the developer is deciding who would be at risk for methicillin-resistant staphylococcus aureus (MRSA) and other serious infections.	The metric focuses on reducing patient risk of developing antibiotic resistance. According to pneumonia guidelines, most patients do not need to receive anti-multi-drug-resistant organism (MRDO) therapy. Approximately 90% of patients who have pneumonia and receive this treatment do not actually need it. Those who do receive the therapy when it is needed are also at risk of side effects, such as kidney injury and even heighted chance of death. The top risk factor for developing a drug-resistant pathogen, such as MRSA, is having had it previously. Another risk factor is exposure to the health care system.
MRSA Percentage: An Advisory Group member commented that they know approximately 51% of the community has MRSA. They asked how the developer accounted for that.	The developer focused on the population that has MRSA growing from their respiratory tract.
Positive Cultures: An Advisory Group member asked how the developer accounted for positive cultures for MRSA or pseudomonas.	Patients with positive cultures for MRSA or pseudomonas qualify for the therapy and are excluded from the measure.
Feasibility: A few Advisory Group members pointed out issues with feasibility including: • That this measure has many exclusions and could potentially be burdensome for hospitals. • That disparate records could make this measure difficult. • That information may not be available at all if a patient is new to a health system or for other reasons related to the electronic medical record (EMR).	In regard to patients going to a different health system or information being missing from the EMR, the measure will miss a small number of patients and that the capture of the data is likely to vary based on how integrated a health system or EMR is. However, overuse is such a prominent issue, that the sensitivity (96%) and specificity (92%) were not affected greatly.

Feedback/Questions	Summary of Developer Response
One Advisory Group member also asked how a facility would be able to identify a specific exclusion criterion: patients with severe community-acquired pneumonia and prior hospitalization with intravenous (IV) antibiotics in the last 3 months.	Prior positive MRSA respiratory or pseudomonas cultures are uncommon, (<1% for MRSA; <2% for pseudomonas), so even if not accurately documented, it is unlikely to impact a hospital's score. (They included that at their three testing sites, the positive MRSA respiratory cultures were represented as: University of Michigan: 0.6%, University of Utah: 0%, VA Healthcare: 0.6%.) Similarly, prior hospitalization with IV antibiotic in the last 3 months (in addition to severe pneumonia) only occurs in 4.3%-10.8% of patients with higher numbers in quaternary hospitals (which are more likely to have EMRs with better data). The goal is not to reach 0% and is, instead, 10%; they emphasized that anything higher than that is overuse. Regarding identifying prior 3 months of hospitalization, test settings were able to manage the request; however, the test settings likely had more integrated health care systems. The developer explored two options: 1) Drop this specific exclusion, which would result in patients being misclassified, or 2) Exclude all patients with severe pneumonia and ignore the additional risk factor of having had exposure to IV antibiotics, which would result in missing a large population that does not qualify for these additional antibiotics. The developer discussed these options with their technical expert panel (TEP). While the exclusion might be difficult or not possible for some hospital systems, it fit best with current guidelines and could potentially improve as more hospital systems become standardized.
Program Inclusion: A few Advisory Group members commented that, given the gaps in the data, the measure may not be appropriate for implementation in a payment program yet.	The developer did not specifically address this comment.
Sepsis and Respiratory Failure: A few Advisory Group members stated that they found the wording of the denominator confusing, believing that the measure would include pneumonia, sepsis, or respiratory failure. A few Advisory Group members also stated that while they understood the developer's upcoding rationale, they struggled with sepsis and respiratory failure being considered uncomplicated. Specific to this measure, one Advisory Group member asked how the developer took into consideration respiratory failure	Please see response for #4540e.

Feedback/Questions	Summary of Developer Response
where patients require higher amounts of oxygen than a simple nasal cannula?	
Hospital At Home: An Advisory Group member asked if the measure includes people who received antibiotics at home as part of hospital at home?	The developer did not include hospital at home. They said they might look at this population in the future.
Value Sets: An Advisory Group member expressed concern that the measure's value sets are not functioning in the matter the developer explained. They highlighted: • That the MRDO value set includes levofloxacin and excludes vancomycin. • That they did not believe the mechanical ventilation exclusion is working. • That the sepsis value set versus the respiratory value set is not accurate. • That the comorbidities value sets are not accurate.	Please see response for #4540e. Regarding vancomycin, IV vancomycin (not oral) is included as anti-MRSA therapy in the clinical quality language (CQL) code—vancomycin is coded separately, as only IV vancomycin should count as MRSA therapy. Both ciprofloxacin and levofloxacin are included in the "BroadSpectrumAntibioticsMDRO" value set as they target *Pseudomonas aeruginosa* and are associated with patient harm and antibiotic resistance.±
De-Escalation: An Advisory Group member pointed out that a treatment plan that is common and consistent with IDSA is to treat individuals who are really sick and have risk factors based on the local environment or have a high risk for MRSA (such as patients who use injection drugs, those with end-stage kidney on hemodialysis, and those who come in with septic shock) with a broader spectrum of antibiotics and de-escalate quickly.	The developer followed current guidelines closely to identify patients with severe pneumonia. The guidelines currently state that if a patient has severe pneumonia based off vital signs, oxygen requirement, and laboratory findings, plus a prior hospitalization with IV antibiotics, then they are eligible for anti-MRSA and anti-pseudomonal coverage. Some of the factors the Advisory Group member highlighted might be risk factors for MRSA but they are not currently included in the guidelines.
ICU Definition: An Advisory Group member asked if ICUs are defined by physical location or ICU level of care?	The ICU is defined by location, which they acknowledged is not perfect.
Black Box Warnings: A patient participant on the committee mentioned that fluoroquinolone has a black box warning, and they believed that patients should be educated on these types of warnings.	The developer agreed wholeheartedly, and this is why they included fluoroquinolones in our anti-pseudomonal antibiotics.
Allergies: An Advisory Group member commented that they did not see allergies accounted for in the exclusions.	The developer did not account for allergies in the measure. Allergies could theoretically allow for anti-pseudomonal therapy to be a prescribed in a patient who—due to allergies—could not tolerate a penicillin or beta-lactam. The issue with attempting to do this was two-fold. First, coding of allergies is poor (so the data would be infeasible to capture) particularly in coding severity of allergy. Second, most patients with penicillin allergies have mild allergies or even intolerances (rather than allergies) and can still receive a beta-lactam. Identifying this in a digital quality metric would not be feasible.

± The developer's full written response can be found in Appendix A.

Key Themes from Advisory Group Feedback, Public Comment, and Staff Assessments

Discussion Categories	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance and Evidence	Advisory Group; Public Comment	Several Advisory Group members highlighted the importance of the measure. Public comments received on this measure further support its importance, expressing that the measure is interoperable and aligns with recommendations from organizations like the CDC, The Joint Commission, and CMS.
Dissenting	Value Sets	Advisory Group	An Advisory Group member stated that several of the value sets were not functioning in the matter in which the developer described.
	Feasibility	Advisory Group	Several Advisory Group members expressed concern over the feasibility of the measure, highlighting its potential burden on hospitals, missing data, and variation in EMRs.
	Validity Testing	Staff Assessment	The validity testing results support a relatively strong inference of validity for the measure, confirming that the measure accurately reflects performance on quality or resource use and can distinguish good from poor performance. However, additional data element testing within at least two EHR vendors is needed.
Mixed	Denominator	Advisory Group	Advisory Group members discussed the inclusion of sepsis and respiratory failure in the measure's denominator. They indicated that while they understood the developer's intention to catch instances of pneumonia where a facility might be "upcoding," it still felt odd to see sepsis and respiratory failure as being considered "uncomplicated." They also added that, as how the denominator is currently worded, they misunderstood that the measure includes 1) pneumonia or 2) sepsis and respiratory failure, believing that individuals could be included for 1) pneumonia, 2) sepsis, or 3) respiratory failure.
	Meaningfulness to Patients	Staff Assessment; Public Comment	Meaningfulness to patients was partially assessed. The information available in the literature was limited and the TEP only included clinicians. However, a public comment shared by Patients for Patient Safety US, an advocacy group, expressed the importance of CBE #4545e and #4540e.
Probing	Equity	Staff Assessment	Though the developer provided some information on inequities, no empirical testing was completed.

CBE 4625e: Emergency Care Capacity and Quality eCQM [Acumen/CMS]

Specifications

Measure Description: *This intermediate outcome eCQM captures the proportion of visits for patients of all ages that experience emergency care access barriers during a one-year performance period.*

Staff Preliminary Assessment Rating

Importance: Met

Rationale: This new measure meets all criteria for "Met" due to its robust evidence-base, clear business case, documented performance gap, significant anticipated impact, well-articulated logic model, and its superiority over existing measures, making it essential for addressing variations in the capacity and quality of ED care.

Feasibility: Met

Rationale: This new measure meets all criteria for "Met" due to its well-documented feasibility assessment, clear and implementable data collection strategy, and transparent handling of licensing and fees, ensuring practical implementation within the healthcare system.

The eCQM Feasibility Scorecard results indicate that the measure is well-supported by the current capabilities of the tested EHR systems in terms of data availability, accuracy, adherence to standards, and workflow integration. These factors collectively ensure that the measure can be implemented effectively and sustainably in a real-world healthcare setting.

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the accountable entity level.

Validity: Met

Rationale: As a new measure, person- or episode-level validity assessment is sufficient.

Going forward, the validity testing results support a weak inference of accountable entity-level validity for the measure, not confirming that the measure accurately reflects performance on quality or resource use and can distinguish good from poor performance.

The stratification methods used are appropriate and demonstrate variation in the prevalence of risk factors across measured entities and show the impact of risk adjustment for providers at high or low extremes of risk.

Equity: Not Met

Rationale: The developer did not address this optional domain.

Use & Usability: Met

Rationale: For initial endorsement, there is a clear plan for use in at least one accountability application, and the measure provides actionable information for improvement.

Public Comment

Number of Comments Received During the Public Comment Period: 30

Comments and their responses from measure developers can be found on the measure page under the "Comments" tab (*Figure 2*).

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Importance: Several Advisory Group members, especially patient participants, expressed that they felt that this was an important measure. They stated that this would be particularly important to the public as long emergency department (ED) stays and limited privacy in hallway beds affect many people. They believed this was an indicator that people would look at before selecting a hospital.	N/A
Actionability: Several Advisory Group members expressed that they had concerns over the actionability of the measure, expressing that wait times in the ED can be a result of many different factors, including hospital location (citing that urban hospitals may have greater populations of unhoused people) or uptick in diseases such as COVID or the seasonal flu.	This measure is going to be a challenge for most hospitals. The intent of the measure is to push hospitals to ensure privacy and safety of patients, such as creating better holding areas.
Subcomponent Reporting: An Advisory Group member asked if the subcomponents to the measure are reported individually. Another Advisory Group member commented that being able to see this data is important for facilities to see where improvement is most needed.	This is a new measure, and reporting has not been established yet. The developer is open to reporting the numerator components and developing a plan with CMS.
Alignment: An Advisory Group member asked how this measure aligns with already existing measures.	OP-18 (Median Time from ED Arrival to ED Departure for Discharged ED Patients) and OP-22 (Patient Left Without Being Seen) overlap with part of the measure's numerator components (*ED length of stay* and *patient left the ED without being evaluated*). Neither OP-18 nor OP-22 are CBE endorsed or an eCQM. The new measure is broader than OP-18 and OP-22, as neither of those measures capture ED boarding or waiting time.±
Transfer Data: An Advisory Group member asked for clarification on how transfer data is used for larger hospitals.	In the Hospital Outpatient Quality Reporting (HOQR) Program, they use inpatient boarding data. For the Rural Emergency Hospital Quality Reporting (REHQR) Program, they use transfer data. Transfer is considered for other components of the measure because the quality and efficiency of the transfer have an impact on patient care.
Exclusions for Psychiatric Patients: A few Advisory Group members asked if the developer had considered an exclusion for psychiatric	The developer does not exclude psychiatric patients but that current measure specifications acknowledge and account for these well-

Feedback/Questions	Summary of Developer Response
patients, pointing out that psychiatric patients often need to be transferred and resources are limited, which is outside the control of the ED.	recognized and unique challenges for patients with psychiatric emergencies. Transfer data are included in the measure.±
Exclusions for Pregnant Patients: An Advisory Group member commented that pregnant patients should also be excluded, as sometimes it may be necessary to deliver a baby in the ED if obstetrics does not have capacity. Another Advisory Group member added that it is important to recognize that in the current political climate, more pregnant people are having difficult accessing care and when pregnancy complications arise, there can be lengthy debate about what kind of care they can access.	The developer appreciated the suggestions regarding exclusions for pregnant people. In their data, pregnancy ED visits are a small amount. They also cited that the Centers for Disease Control and Prevention's (CDC's) National Hospital Ambulatory Care survey reported a national estimate of 0.0%. If concerns for this population continue to arise, the developer said they can explore potential exclusions with CMS, which they anticipate would have limited impact on the measure testing results because the rate of occurrence is very low.
Evidence: An Advisory Group member pointed out that the evidence backing that a patient should be within a treatment room within 1 hour is not strong. They stated that the developer cited two studies to support this concept, with one being based in Canada and the other being time to triage. Several Advisory Group members agreed that the time should be time to triage and not to treatment room.	The developer confirmed the evidence supports arrival to triage. The intention of numerator criteria #1 is to ensure that patients, in a timely fashion, are placed in a treatment room or an area with privacy during history-taking and physical examination, rather than in a hallway or an ED lobby. In the routine order of care in an ED, triage is conventionally completed upon patient arrival, so it would be assumed to occur within numerator criteria #1's 1-hour timeframe. This 1-hour threshold was vetted by a TEP and aligns with guidance from the American College of Emergency Physicians. CMS will continue to monitor whether this 1-hour definition remains reasonable. The developer will continue to explore other avenues to better capture time to triage and advocate for better standards to capture triage times more precisely.
Triage: Several Advisory Group members said the measure needs to account for the severity of illness or injury (such as through the Emergency Severity Index [ESI]) that a patient comes into the ER with. An Advisory Group member emphasized that it is not safe or appropriate for all patients to be placed within a treatment bed. One Advisory Group member asked if it would be possible for another measure to look at the criticality of patients.	The ECCQ measure is not stratified or adjusted for differences in severity of illness distribution because each of the numerator outcomes measured should be considered an access failure regardless of illness severity. The developer's TEP vetted this approach and considered it agreeable. Using a triage score-based adjustment is not feasible because of a lack of uniform triage scoring due to the increasing use of new triage acuity scores and tools. Prior literature has suggested that ESI scoring in ED triage may be prone to bias as well as very low reliability given

Feedback/Questions	Summary of Developer Response
	high inter- and intra-rater variability (Essa CD, Victor G, Khan SF, Ally H, Khan AS. Cognitive biases regarding utilization of emergency severity index among emergency nurses. Am J Emerg Med. 2023 Nov;73:63-68. doi: 10.1016/j.ajem.2023.08.021. Epub 2023 Aug 12. PMID: 37619444).
Patient Choice: An Advisory Group member asked how the developer considered patient choice when it comes to selecting and ER, and that patients may cross county lines and bypass other hospitals to go to a certain ER.	The developer discussed patient selection. The measure is intended to allow the comparison of similar ERs to one another, which is meant to neutralize some of the effect from patient selection.
Unintended Consequences: Several Advisory Group members expressed that they were concerned the measure, as currently specified, would result in unintended consequences. They discussed that this measure may hurt safety-net hospitals and city- and state-run hospitals; that "drive-bys" (patients avoiding some EDs while flooding others) may become more frequent; that patients may be pushed out of ERs; that ERs may be closed altogether when at capacity; patients may be admitted or placed in observation when not appropriate; and that not taking into account triage level will compromise the safety of more severe patients and interfere with the flow of the ED.	This measure is going to be challenge for most hospitals. The intent of the measure is to push hospitals to ensure privacy and safety of patients, such as creating better holding areas. The intent the measure to be pay-for-reporting, so there would be no financial consequences on the hospital. Instead, strong performance will likely result in cost savings. CMS will evaluate for such unintended consequences during measure implementation and conduct impact and surveillance analyses by hospital characteristics to ensure that the measure does not include any systematic biases. Regarding consequences for hospitals with specific characteristics (such as safety-net hospitals or city and state hospitals), the HOQR version of the measure utilizes volume standardization to allow hospitals of similar characteristics to be compared to one another. Additionally, the version of the measure being considered for the REHQR program allows REHs to only be compared to other REHs. Regarding "drive-bys," the developer did not expect this to occur and that there is little evidence. However, they will continue monitoring the impact of the measure on neighboring hospitals. Regarding observation status, the developer had that conversation with their TEP and believed that clinicians would continue to use that status in a clinically relevant manner and not game the measure. If that is a continued concern, they could add another stratification. ±
Volume Standardization: An Advisory Group member stated that, based on the submission materials, they were unclear of if volume standardization is an intrinsic part of the measure or not. The Advisory	Volume standardization is implemented in the HOQR version but not in the REHQR version. The volume standardization is intended to compare hospitals with similar characteristics. The testing data

Feedback/Questions	Summary of Developer Response
Group member also added that it is not preferable to stratify to hospital characteristics that change over time, such as ED volume.	showed that similar hospitals tend to have similar volumes. The developer proposed bands of 20,000 ED visits per year. This approach makes it straightforward to adjust without having to list numerous risk adjustors. Based on their data, EDs were unlikely to cross the 20,000 threshold per year. They stated that they will continue to monitor the threshold to see if it remains appropriate and can adjust it as needed.
Stratification by Race: An Advisory Group member asked for clarification on whether the developer is proposing the measure be stratified by race.	The developer considered race but are not stratifying by it at this time. They will discuss the issue further with CMS.
Stratification by Age: An Advisory Group member asked for clarification on how the measure is stratified by age, stating that the stratification criteria are broken out by two age groups (under 18 and over 18) but the discussion mentioned three age groups.	The developer confirmed the measure is stratified by two age groups: under 18 and over 18.
Baseline Data: An Advisory Group member asked for information about the baseline performance data for the measure.	The developer used three different data sets from 2022-2023, during which 32 entities captured over 2 million encounters. One data set spans many health care systems and includes four rural EDs.
Staffing: An Advisory Group member inquired whether there is another quality measure being developed to address ED staffing for patients who are boarded.	Currently, there are no existing quality measures in the HOQR or REHQR programs that assess ED staffing for patients who are boarding. CMS will take this comment into consideration.
Pandemic or Other Major Events: An Advisory Group member asked if the measure accounts for another pandemic or other major trauma event.	The developer will defer to existing and future CMS exemptions and exceptions policies as well as public-reporting freezes in the event of a future public health emergency. Based on data from the last public health emergency, the developer expects such an event to impact all hospitals equally. Finally, this measure is being considered for pay-for-reporting programs, so facilities will not be financially impacted based on their measure performance.
Quality of Assessment: An Advisory Group member asked if the developer considered the quality of the assessment that the patient received in the ED.	The closest assessment they have is the data element validity, which they conducted a chart review on over 250 patient charges and compared those to the EHR associated with the treatment room and found a good match.

± The developer's full written response can be found in Appendix A.

Key Themes from Advisory Group Feedback, Public Comment, and Staff Assessments

Discussion Categories	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance	Advisory Group; Public Comment	Several Advisory Group members, particularly the patient participants, and members of the public highlighted that this is an important measure to the public, as many individuals are affected by lengthy wait times in the ER.
Dissenting	Actionability	Advisory Group	Several Advisory Group members expressed concern over the feasibility of the measure, given that many factors that contribute to lengthy ED wait times are complex and outside the control of the facility.
	Exclusions	Advisory Group; Public Comment	The Advisory Group discussed whether additional exclusions should be applied to the measure, focusing predominantly on psychiatric patients and pregnant patients. In a public comment, the American College of Emergency Physicians (ACEP) requested the exclusion of ED visits with a transfer out status from the measure to ensure accurate reflection of care quality, as rural hospitals often face challenges in transferring patients, and larger hospitals bear the responsibility for coordinating such transfers.
	Evidence	Advisory Group	Several Advisory Group members voiced concerns with the expectation that all patients be placed within a treatment room within 1 hour, stating that this is not backed by evidence.
	Triage	Advisory Group	Several Advisory Group members stated that the measure should consider the patient's severity, such as through using ESI.
	Unintended Consequences	Advisory Group	Several Advisory Group members emphasized their concern that the measure would result in unintended consequences that are contrary to the purpose of the measure and would possibly jeopardize the safety of some patients.
Mixed	Measure Specifications and Applicability	Public Comment	The ACEP provided feedback on specific outcomes, such as the need for privacy in treatment areas and clarity on patients leaving without evaluation. They support a 4-hour maximum boarding time and suggest future targets for shorter durations. ACEP supports the stratification and suggests further age group stratification (18-65 and 65 years and older). They recommend reporting without volume standardization for certain outcomes. ACEP suggests structuring the measure as a composite measure, with boarding weighted more heavily. They support its inclusion in the Hospital OQR Program but not in the Rural Emergency Hospital Quality Reporting Program due to potential data skewing factors.

CBE 4700e: Rate of Timely Follow-up on Abnormal Screening Mammograms for Breast Cancer Detection [Brigham and Women's Hospital]

Specifications

Measure Description: *This electronic Clinical Quality Measure (eCQM) reports the percentage of female patients aged 40 to 75 years with at least one abnormal screening (BI-RADS 0) or screening-to-diagnostic (BI-RADS 4, 5) mammogram during the measurement period (i.e., calendar year) who received timely diagnostic resolution defined as either follow-up imaging with negative/benign/probably benign results or a breast biopsy within 60 days after their index (i.e., first) abnormal screening mammogram.*

Negative/benign/probably benign follow-up imaging was defined as diagnostic mammography, breast ultrasound or magnetic resonance imaging (MRI) with BI-RADS ratings of 1, 2, or 3. Relevant diagnostic breast biopsy procedures were defined as core needle biopsy, fine needle aspiration, and surgical excision.

Breast Imaging – Reporting and Data System (BI-RADS) ratings: 0-incomplete, 1-negative, 2-benign, 3-probably benign, 4-suspicious, 5-highly suggestive of malignancy.

Staff Preliminary Assessment Rating

Importance: Not Met but Addressable

Rationale: This eCQM process measure for breast cancer screening and diagnostics is designed to fill existing gaps in quality assessment by enabling facilities to monitor the timeliness and completeness of care. Evidence supports that mammographic screening has significantly reduced breast cancer mortality since the 1990s and that early detection can substantially lower treatment costs. However, challenges remain, particularly for racial and ethnic minorities and low-income groups who are more likely to experience delays in follow-up after abnormal screenings, potentially leading to worse outcomes. This measure, alongside regular quality assessments and interventions such as patient navigation and EHR reminders, aims to enhance follow-up rates and promote health equity. Patient input has underscored the importance of this measure. The main limitation is the developer's logic model does not clearly depict what inputs and activities are needed to report and improve on this metric.

Feasibility: Met

Rationale: This eCQM leverages data that are routinely collected across various health systems and facility groups, all in electronic format, which facilitates the implementation of the measure. The developer identified challenges with inconsistent use of structured fields for BI-RADS. This prompted the development of a string search method to extract these results from unstructured EHR fields, which showed near-perfect performance in initial applications. These findings did not change the final measure specifications.

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the accountable entity level.

Validity: Not Met but Addressable

Rationale: The validity rating is based on patient-/episode-level (data element) testing only, which is acceptable for a new eCQM. The accountable entity level testing, while not required, was not sufficient. For face validity, a larger TEP of at least 12 members including patient representatives and broad representation from potential measure users is preferred. In addition, a Likert scale of at least five responses is preferred to demonstrate consensus. The ICC should be reported under reliability, not validity. Further testing with additional sites and within at least two EHR vendors is necessary for the future.

Equity: Not Met but Addressable

Rationale: The eCQM performance rates, stratified by demographics such as age, race, ethnic group, primary insurance, and primary language, revealed significant disparities; white and English-speaking patients were more likely to achieve diagnostic resolution within 60 days after an abnormal screening mammogram. Despite no significant differences found across the six facility groups due to smaller sample sizes, further analyses are planned to address these disparities, particularly following lower performance rates observed in 2023. The developer could also provide additional information in the submission itself describing methods and exploring the interpretation of the disparities findings and how they might be used to improve health care.

Use & Usability: Met

Rationale: The new measure, intended for public reporting and quality improvement, addresses challenges in tracking follow-up data after mammographic screenings due to resource limitations. Entities are encouraged to standardize policies for timeliness and reporting, enhance staff training on data systems, and improve EHR interoperability to ensure consistent data tracking. These steps aim to facilitate targeted interventions to improve timely diagnostic resolutions.

Public Comment

Number of Comments Received During the Public Comment Period: 1

Comments and their responses from measure developers can be found on the measure page under the "Comments" tab (*Figure 2*).

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Follow-up Care: Several Advisory Group members asked for clarification on what would happen if a patient moved to a different health system for a second opinion or follow-up care, particularly if the systems are not using the same electronic medical record (EMR) system. They asked about potential data issues if a patient did not receive care at the same location. Another Advisory Group member stated that it would be helpful to have data on how often patients switch providers between an initial screening and next steps.	The developer advocates for the patient to receive second opinions. This issue is not unique to their measure but affects all electronic clinical quality measures. Patients tend to seek second opinions more commonly on diagnostic imaging and not the initial screening. The developer found rates of out-of-system diagnostic resolution to be low, considering the high performance of all three health systems on this eCQM. They said they applied a benchmark of 90% to accommodate for out-of-system follow-ups for the performance gap assessment.[±]
Importance: Several Advisory Group members expressed that this was an important measure, emphasizing that early detection is vital in treating breast cancer.	N/A.
Timeframe: Several Advisory Group members asked for clarification about what takes place within the 60 days specified in the measure. Several Advisory Group members stated a quick follow-up time helps to alleviate patient anxiety. Others mentioned that rural and under-resourced facilities might require the time allotted in the measure or even longer. One Advisory Group member highlighted that while they personally prefer the 60-day timeframe, this measure does not currently align with a National Committee for Quality Assurance (NCQA) measure, Follow-Up after Abnormal Breast Cancer Assessment (BCF-E).	During the 60 days, the patient should have had repeat imaging and, if needed, a biopsy. The 60 days is not intended to be the timeframe in which the patient receives the results of their initial screening. The developer had in-depth conversations regarding a 60-day or 90-day timeframe, which were both appropriate in current literature. After 90 days, there is an important difference in terms of later-stage diagnoses. In part, the developer took guidance from the National Breast and Cervical Cancer Detection Program, which serves uninsured or underinsured women. In addition, their TEP recommended the 60-day timeframe because of clinical impact and also to alleviate the stress and anxiety that patients experience while waiting. The developer said they have not had any conversations around 30 days.

Feedback/Questions	Summary of Developer Response
	The mentioned NCQA measure is not 90 days total but 90 days between each step. As this measure is 60 days for the whole cycle, it is a significantly shorter timeframe.±
Value: An Advisory Group member asked what value the measure would create in addition to measures that the NCQA has proposed for the Healthcare Effectiveness Data and Information Set (HEDIS): Documented BI-RADS Assessment after Mammogram (DBM-E) and Follow-Up after Abnormal Breast Cancer Assessment (BCF-E).	The measure is patient centered and focuses on the timeliness of diagnostic resolution within 60 days after an abnormal (i.e., inconclusive) screening mammogram. The measure is more comprehensive by also including breast MRI as appropriate diagnostic follow-up. The BCF-E HEDIS measure is based on mammography episodes, so a patient may be included in the denominator multiple times in one measurement year. As noted above, the measure allows for up to 180 days (90 days after inconclusive screening mammogram plus 90 days after high-risk assessment on diagnostic follow-up imaging) to diagnostic resolution requiring biopsy.±
Cost: A few Advisory Group members expressed concern about how cost may play into the measure and asked what would be covered for a patient beyond the initial mammogram and if there could be misconceptions about the cost of additional testing.	Follow-up imaging is often associated with high out-of-pocket costs. There has been a recent push for insurance companies to cover the full cycle. The measure provides the opportunity to observe differences between groups and potentially influence further policies to cover costs associated with the full breast cancer screening process, including breast biopsy.±
False Positives: The committee discussed how false positives may affect the follow-up rate: • A few Advisory Group members stated that patients may be less likely to do follow-up if they routinely receive false positives, with several Advisory Group members acknowledged that this decision is a personal choice. • A few Advisory Group members, particularly patient participants, commented that they would rather go through the whole process and receive a negative response to alleviate worry and that follow-up is important in early detection. • One Advisory Group member stated that false positives are well incorporated into the measure's logic.	Patients are less likely to follow up if they have had false positives in the past. There is a need to respect patient wishes around follow-up, but the purpose of the measure is to help detect breast cancer as quickly as possible and move forward with timely intervention when necessary. The developer has had recommendations stating not to expect 100% compliance with the measure, so they will continue to talk about a lower benchmark to allow flexibility for these types of cases.
Terminology: An Advisory Group member commented on how the measure refers to Bread Imaging Reporting and Database System (BI-RADS) scores of 4 and 5 as "diagnostic." They said that as the	The developer is open to feedback on hoping they could be clearer in their terminology. The language they currently use is reflective of what they have seen across the health care systems. They measure is

Feedback/Questions	Summary of Developer Response
measure focuses on initial screenings, it would be more accurate to call these scores "highly suggestive."	trying to capture individuals who come in for something suspicious and go straight to a diagnostic mammogram.
Feasibility: An Advisory Group member asked if all sites have data fields that are discrete enough to capture final diagnosis or if that information is embedded in free text.	EPIC and Cerner do have discrete fields for the final BI-RADS rating for both the initial screening mammogram and diagnostic imaging. However, facilities do not always use the structured fields. While their recommendation is that facilities do use those, they have also created a string-search algorithm to pull that information when it is not in the structured field. In a random sample, the algorithm had near perfect performance. Therefore, even when the information is not captured in fields, it still is straightforward to pull the needed information and calculate the measure.±
Age: A few Advisory Group members expressed that individuals need breast cancer screenings past the age of 75, highlighting that this is a critical health issue for many Americans and that people are living longer. One Advisory Group member pointed out that the clinical standard is to perform screening mammograms on women ages 40-75; they said that it is possible a different measure may be needed for diagnostic mammograms in patients with a prior history of cancer or other factors.	The developer appreciated the comment and that they are aligned with universal standard guidelines at this point.
Proprietary Component: An Advisory Group member pointed out that the biopsy value set is limited to Current Procedural Terminology (CPT), which is proprietary. The Advisory Group member suggested the developer may consider broadening beyond CPT.	The developer did not specifically address this comment.
Standalone Breast Centers: An Advisory Group member asked if standalone breast centers would also report on the measure.	The developer did not specifically address this comment.
Equity: An Advisory Group member noted that this measure could promote equity for minority populations by aligning with guidelines that lower the recommended screening age to 40, thereby improving care for these groups.	The developer did not specifically address this comment.
Scientific Acceptability: A committee member asked why split-sample Spearman's Rank Correlation Coefficients and Interclass Correlation Coefficients (ICC) were presented as validity rather than reliability.	These analyses were placed in the wrong section in error and should be considered reliability testing.

± The developer's full written response can be found in Appendix A.

Key Themes from Advisory Group Feedback, Public Comment, and Staff Assessments

Discussion Categories	Key Themes	Source of Comment	Summary of Comments
Dissenting	Feasibility	Advisory Group; Public Comment	The Advisory Group expressed concern about how easy the measure would be to track if an individual goes to a different health care system for follow-up care or a second opinion, particularly if the health care systems do not use the same EMR. The Advisory Group further recognized that not all sites may use structured fields or have data fields that are discrete enough to capture final diagnosis (e.g., BI-RADS). This was also identified in a public comment submitted by the American Medical Association (AMA).
	Testing	Advisory Group; Staff Assessment; Public Comment	The validity rating is based on patient-/episode-level (data element) testing only, which is acceptable for a new eCQM. However, only one EHR vendor was used for validity testing, which was also noted in a public comment submitted by the AMA. In addition, the ICC should be reported under reliability, not validity. Further testing with additional sites is necessary for the future.
	Timeframe	Advisory Group	The Advisory Group discussed at length whether the 60-day timeframe is appropriate. Most of the Advisory Group agreed that it was, with consideration given to patient anxiety during wait times, capabilities of rural and under-resourced facilities, and alignment with other measures.
	Age	Advisory Group	A few patient participants of the Advisory Group expressed that women over the age of 75 still need to have mammograms. Another member of the committee emphasized that the measure follows current guidelines and that another measure may be needed to address the needs the other Advisory Group members highlighted.
Mixed	Importance	Advisory Group; Staff Assessment	The Advisory Group highlighted that this is an important measure, given that breast cancer is a treatable cancer, and that early detection is key. The developer's logic model does not clearly depict what inputs and activities are needed to report and improve on this metric.
Probing	Equity	Staff Assessment	The analytic approach and interpretation of results were not specified in the submission.

CBE 4705e: Rate of Timely Follow-up on Positive Stool-based Tests for Colorectal Cancer Detection [Brigham and Women's Hospital]

Specifications

Measure Description: *This electronic Clinical Quality Measure (eCQM) reports the percentage of patients aged 45 to 75 years with at least one positive stool-based colorectal cancer screening test (i.e., high-sensitivity guaiac fecal occult blood test, fecal immunochemical test, or Cologuard) during the measurement period (i.e., calendar year) who completed a colonoscopy within 180 days after their index (i.e., first) positive stool-based test result date.*

Staff Preliminary Assessment Rating

Importance: Not Met but Addressable

Rationale: This new eCQM process measure for colorectal cancer screening is designed to address significant gaps in the screening process by enabling facilities to monitor and enhance the timeliness of follow-up colonoscopies after positive stool-based tests. The developer provides evidence of effective interventions that can overcome the challenges to improving follow-up rates. The 180-day follow-up period for colonoscopies, recommended by the American Gastroenterological Association, lacks grading in the submission, and the logic model focuses more on measure development than on specifying necessary actions for entities to achieve this measure.

Feasibility: Met

Rationale: For this eCQM all necessary data elements are consistently collected in electronic format. The feasibility assessments conducted across these facilities yielded consistent feasibility scores, demonstrating that the structured data fields used did not influence the final specifications of the measure. The measure's non-proprietary nature ensures broad accessibility.

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the accountable entity level.

Validity: Not Met but Addressable

Rationale: The validity rating is based on patient-/episode-level (data element) testing only, which is acceptable for a new eCQM. The accountable entity level testing, while not required, was not sufficient. For face validity, a larger TEP of at least 12 members including patient representatives and broad representation from potential measure users is preferred. In addition, a Likert scale of at least five responses is preferred to demonstrate consensus. The ICC should be reported under reliability, not validity. Further testing with additional sites and within at least two EHR vendors is necessary for the future.

Equity: Not Met but Addressable

Rationale: The eCQM performance rates, stratified by demographic and clinical factors such as age, sex, race, and type of stool-based test, highlighted significant differences by primary insurance and test type, and further analyses across six facility groups revealed significant disparities in timely colonoscopy rates influenced by ethnic group and primary language. The developer could also provide additional information in the submission itself describing methods

and exploring the interpretation of the disparities findings and how they might be used to improve health care.

Use & Usability: Met

Rationale: The new measure, intended for public reporting and quality improvement, aims to improve timely follow-up colonoscopy rates within 180 days of a positive stool-based test by reducing site-related barriers and implementing standard protocols. Evidence-based interventions such as patient navigation, case management, patient education on bowel preparation, timely communication of results, and EHR reminders have shown to decrease time to follow-up.

Public Comment

Number of Comments Received During the Public Comment Period: 3

Comments and their responses from measure developers can be found on the measure page under the "Comments" tab (*Figure 2*).

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Age: An Advisory Group member commented that they do not like the age cap on this measure.	The developer did not specifically address this comment.
Timeframe: An Advisory Group member commented that they believed 180 days is too long. Another Advisory Group member stated that 180 days might not be long enough, given the work clinicians may need to do to convince patients to actually undergo the colonoscopy.	The developer shared how they selected the 180-day timeframe. They did not find guidance on a timeline, so they followed the recommendation of their TEP to look at the literature and find the shortest amount of time that leads to clinical outcomes. They found that after 180 days, there is a higher likelihood of being diagnosed with a more advanced cancer; however, follow-up is still incredibly low after 180 days. Studies indicated that, across 39 health systems, only about 50% of patients received a colonoscopy after a positive stool-based test. Therefore, they selected 180 days because they felt the timeline was not too short to unmotivated facilities and they knew it would still influence clinical outcomes and improve rates. As the rates improve, they can modify the measure to make the follow-up period shorter.
Inclusion: An Advisory Group member requested clarification on the inclusion criteria for the measure.	This measure captures patients who prefer not to have a colonoscopy and have a fecal immunochemical test (FIT) that has a positive result and then require a follow-up colonoscopy.
Patient Reluctancy: A few Advisory Group members stated that this measure may be difficult because patients who opt not to go for a colonoscopy in the first place are often reluctant even after a FIT. They suggested the developer allow exclusions for patients who decline.	The developer recognized that it is difficult to have patients follow up with a colonoscopy, thus the intent of this measure; however, they highlighted that this is part of the conversation that needs to happen between clinicians and patients so that patients understand the importance of follow-up. The developer also added that excluding patients who decline a colonoscopy could put the measure at risk for gaming.±
Consideration of Care Coordination within the Measure: A few Advisory Group members stated that a primary care physician usually (PCP) initiates the stool test but a specialist would perform the colonoscopy. They asked for more information regarding what happens if the specialist and the PCP are in different systems, how	The developer stated that they know that in about 10-15% of cases, a patient may be referred outside of their system. The noted that the American Gastroenterological Association (AGA) recommends that 95% of patients have a colonoscopy after a positive stool-based test. To accommodate the patients who may go outside of system, the developer dropped the benchmark to 80%.

Feedback/Questions	Summary of Developer Response
often this occurs, and for the developer to comment on issues related to communications and follow-up bias.	They noted that, after discussing this with their TEP, they also opted for an integrated delivery system at the facility group level, meaning that PCPs and gastrointestinal (GI) specialists who serve a specific region or catchment area are considered to be within the same facility group. The measure works by extracting any colonoscopy completed within the integrated delivery system. They added that approximately 90% of the time a patient stays within the same facility group.±
Self-Referral Incentives: An Advisory Group member asked, if based on the developer's explanation of how they define facility groups, if a provider would be incentivized to self-refer within their group.	The developer stated that providers are incentivized to refer within the integrated delivery system/facility group that serves their patient population.
Standalone and Rural Practices: A few Advisory Group members asked for clarification on if, based on how the measure defines facility groups, standalone practices would be excluded from the measure. They expressed concern that a considerable proportion of rural care is still provided by individual physician practices that would be excluded from the measure. They suggested reconsidering the stratification of rural and urban.	The developer also clarified that standalone primary care clinicians are excluded from the measure and that only practices that are part of a larger system can appropriately use the measure at this time. The developer said they will consider the stratification.
Hybrid Measure: An Advisory Group member recommended that the developer consider making this a hybrid measure where they leverage Logical Observation Identifies Names and Codes (LOINC) to track follow-up care, regardless of where it is provided.	The developer thanked the commenter for the recommendation of a hybrid measure and stated that they would look into that.
Unrepresented Testing: An Advisory Group member commented that the measure used a limited number of testing sites and that, across the country, it seems as if the measure will not be applicable in a large proportion of places.	The developer stated that this is a new measure, and there are limitations to the data they can present as of now. However, they also stated they are working with additional health facility groups that will be added to the analyses.
Scientific Acceptability (i.e., Reliability and Validity): Advisory Group members had several questions related to the reliability and validity testing results: • **Face Validity:** An Advisory Group member commented that this topic has enough literature that the developer could be looking at other forms of validity. • **Split-Half Testing for Validity:** An Advisory Group member stated that they found it odd that the developer seemed to be placing split-half reliability under validity. • **Signal-to-Noise Reliability Testing:** An Advisory Group member commented that they found Table 4 to be confusing. They highlighted that the developer had specified the measure	Regarding face validity testing, Battelle stated that face validity of the performance score is acceptable for new measures, and this submission meets the minimum of what is expected for a new measure. They added that the developer could keep this comment in mind for maintenance if the measure moves forward. The developer also thanked the commenter for suggestions of other types of validity they can investigate. With respect to the split-half testing, the developer stated the analyses were placed in the wrong section in error and should be considered reliability testing.

Feedback/Questions	Summary of Developer Response
for annual reporting, but that median signal-to-noise ratio (SNR) varies from year to year in a way that is odd.	Lastly, for the signal-to-noise testing, the developer stated that they opted to report the measure annually because there have been so many factors that affect access to care, including COVID and stool-based testing becoming recommended, and they wanted to reflect that. They acknowledged that more than 55% of their data is from 2022-2023, which is why those results are better and more reliable.
Cost: An Advisory Group member pointed out that once a patient has a positive stool-based test, the colonoscopy would become diagnostic. They expressed concern that this would be an out-of-pocket cost for many patients. Another Advisory Group member mentioned that even with the new policy changes noted by the developer, they were not sure how consistently ICD-10 codes were being appropriately used on claims, which are needed to avoid the out-of-pocket costs.	The developer stated that as of January 2023, a diagnostic colonoscopy should be covered by the insurance provider, although they acknowledged that bowel prep still has significant out-of-pocket costs. The developer added that while out-of-pocket costs may affect the likelihood of missing or delaying a follow-up colonoscopy, the measure provides the opportunity to observe differences between groups and potentially influence policy to cover costs associated with the full colorectal cancer-screening process. ±
Colonoscopy Attempts: An Advisory Group member asked how health systems would report attempts to complete a colonoscopy where the patient could not tolerate the bowel prep, had inadequate prep, or had poor visualization of the bowel.	The developer stated that the measure, as currently conceptualized, counts if the colonoscopy was completed even if the bowel prep or visualization was not ideal.

± The developer's full written response can be found in Appendix A.

Key Themes from Advisory Group Feedback, Public Comment, and Staff Assessments

Discussion Categories	Key Themes	Source of Comment	Summary of Comments
Dissenting	Generalizability	Advisory Group; Public Comment	Several members of the Advisory Group expressed concern over how easily this measure could be applied across the country, with some members pointing out that many rural practices could not use this measure as it is currently configured and one Advisory Group member pointing out that the testing sites were not representative. An Advisory Group member suggested that a hybrid measure would alleviate some of these issues. In a public comment, the AMA shared concern with the measure's inability to account for follow-up care received at different facilities, which may lead to inaccurate reflections of a facility's performance. However, the ACG and ASGE recommend including exceptions for follow-up colonoscopies conducted outside the system to avoid penalizing facilities unfairly.

Discussion Categories	Key Themes	Source of Comment	Summary of Comments
	Inclusion of Positive Blood Tests	Public Comment	Several public comments expressed that while the U.S. Preventive Services Task Force does not currently include blood-based tests among its recommendations for methods for screening for colorectal cancer (CRC), patients with positive blood-based CRC screening tests should be included in the measure as they require a follow-up colonoscopy.
	Testing	Advisory Group; Staff Assessment; Public Comment	An Advisory Group member commented that the reliability and validity information were confusing as presented and that some of the information under validity should have been considered reliability. Specifically, the staff assessment noted that the ICC should be reported under reliability, not validity. Further testing with additional sites is necessary for the future. However, the validity rating is based on patient-/episode-level (data element) testing only, which is acceptable for a new eCQM. Public comments expressed the need for expanded testing, noting that current evaluation criteria require developers to complete validity testing on at least two EHR systems, but only one system was used for the data element validity analyses.
Mixed	Importance	Advisory Group; Staff Assessment; Public Comment	The Advisory Group disagreed over whether 180 days is an appropriate timeframe. One Advisory Group member stated they felt it was too long while another stated clinicians might need even longer to convince patients to return for a colonoscopy. The 180-day follow-up period for colonoscopies, recommended by the American Gastroenterological Association, lacks grading in the submission, and the logic model focuses more on measure development than on specifying necessary actions for entities to achieve this measure. However, the developer notes that reducing site-related barriers and implementing standard clinical workflows, along with interventions like patient navigation and EHR reminders, can significantly improve timely follow-up rates. In a public comment, the American College of Gastroenterology (ACG) and American Society for Gastrointestinal Endoscopy (ASGE) suggested adjusting performance targets to be evidence-based rather than set at ≥80%.
	Patient Reluctancy	Advisory Group	A few members of the Advisory Group expressed that they felt the measure would be difficult to implement because patients are reluctant to return for colonoscopies.
Probing	Equity	Staff Assessment	The analytic approach and interpretation of results were not specified in the submission.

Appendix A

Following the Advisory Group meeting, developers/stewards had the opportunity to provide further written responses to feedback and questions from Advisory Group members. An abridged summary of these additional responses is presented in the discussion guide tables. The complete responses from developers/stewards, edited by Battelle staff for clarity and grammatical correctness, are included below.

CBE #4625e: Full Responses Written by the Developer

Feedback/Questions	Written Developer Response
Alignment: An Advisory Group member asked how this measure aligns with already existing measures.	Two existing measures overlap with the ECCQ measure's *ED length of stay* and *patient left the ED without being evaluated* numerator components. These measures include Median Time from ED Arrival to ED Departure for Discharged ED Patients (OP-18) in CMS's HOQR and REHQR programs and Left Without Being Seen (OP-22) in CMS's HOQR Program. Neither OP-18 nor OP-22 are CBE endorsed. The ECCQ measure's outcome is broader than both OP-18 and OP-22, neither of which capture ED boarding, which is a critical component of the ECCQ measure. ED boarding is not captured by any other currently publicly reported measure. The ECCQ measure also captures the outcome of waiting time (time from arrival to first treatment space with audiovisual privacy), which is also not captured by any currently publicly reported measure. The outcome calculation for OP-18 is based on the median, whereas the overlapping ECCQ eCQM component is based on a threshold of 8 hours. Use of a median can mask poor performance when the distribution is skewed to the right. Furthermore, by capturing four different components in one measure, the ECCQ eCQM provides a window into broad aspects of ED quality access gaps with one measure. In addition, neither OP-18 nor OP-22 are eCQMs; therefore, this ECCQ measure improves upon these existing measures as the ECCQ measure is an eCQM, and hospitals will receive data around the individual components of the ECCQ eCQM.
Exclusions for Psychiatric Patients: A few Advisory Group members asked if the developer had considered an exclusion for psychiatric	In terms of the measure's specifications toward patients needing psychiatric care, current measure specifications acknowledge and

Feedback/Questions	Written Developer Response
patients, pointing out that psychiatric patients often need to be transferred and resources are limited, which is outside the control of the ED.	account for these well-recognized and unique challenges for patients with psychiatric emergencies. Specifically, the stratified nature of measurement allows CMS and hospitals to review a measure score for patients with psychiatric emergencies without necessitating that publicly reported scores include such patients for whom attribution of quality may be less closely linked to the hospital. Psychiatric emergencies capture an important population that many hospitals face challenges in serving. Therefore, the stratifications of this measure provide a dual benefit of (i) allowing hospitals to pinpoint areas for improvement and (ii) recognizing hospital efforts in areas that they are performing well that would not be visible without stratifications.
Unintended Consequences: Several Advisory Group members expressed that they were concerned the measure, as currently specified, would result in unintended consequences. They discussed that this measure may hurt safety-net hospitals and city- and state-run hospitals; that "drive-bys" (patients avoiding some EDs while flooding others) may become more frequent; that patients may be pushed out of ERs; that ERs may be closed altogether when at capacity; patients may be admitted or placed in observation when not appropriate; and that not taking into account triage level will compromise the safety of more severe patients and interfere with the flow of the ED.	Regarding unintended consequences for hospitals with specific characteristics, such as safety-net hospitals or city/state hospitals, we highlight that the HOQR version of the measure utilizes volume standardization to allow hospitals of similar characteristics to be compared to one another. Additionally, the version of the measure being considered for the REHQR program allows REHs to only be compared to other REHs. CMS will evaluate for any unintended consequences during measure implementation and conduct impact and surveillance analyses by hospital characteristics to ensure that the measure does not include any systematic biases. In terms of concerns regarding increased occurrences of "drive-bys," we do not expect that this measure will result in such an unintended consequence as there is little evidence of this. However, we can continue monitoring the impact of ECCQ on neighboring hospitals. Generally, we highlight that the programs in which the ECCQ measure is being considered are pay-for-reporting; therefore, CMS will not financially penalize facilities based on measure performance. We emphasize that strong performance on this measure will likely result in cost savings, as meeting numerator criteria may result in hospital revenue reduction and increased costs associated with length of stay. (See associated references for more information.) References: Baloescu, Cristiana, Jeremiah Kinsman, Shashank Ravi, Vivek Parwani, Rohit B. Sangal, Andrew Ulrich, and Arjun K. Venkatesh.

Feedback/Questions	Written Developer Response
	2021. "The Cost of Waiting: Association of ED Boarding with Hospitalization Costs." The American Journal of Emergency Medicine 40 (February 2021): 169–72. https://doi.org/10.1016/j.ajem.2020.10.058. Dyas, Sheila R., Eric Greenfield, Sherri Messimer, Swati Thotakura, Sampson Gholston, Tracy Doughty, Mary Hays, Richard Ivey, Joseph Spalding, and Robin Phillips. 2015. "Process-Improvement Cost Model for the Emergency Department." Journal of Healthcare Management 60 (6): 442–57. https://doi.org/10.1097/00115514-201511000-00011. Pines, Jesse M., Robert J. Batt, Joshua A. Hilton, and Christian Terwiesch. 2011. "The Financial Consequences of Lost Demand and Reducing Boarding in Hospital Emergency Departments." Annals of Emergency Medicine 58 (4): 331–40. https://doi.org/10.1016/j.annemergmed.2011.03.004. Schreyer, Kraftin E., and Richard Martin. 2017. "The Economics of an Admissions Holding Unit." Western Journal of Emergency Medicine 18 (4): 553–58. https://doi.org/10.5811/westjem.2017.4.32740.

CBE #4545e: Full Responses Written by the Developer

Feedback/Questions	Written Developer Response
Value Sets: An Advisory Group member expressed concern that the measure's value sets are not functioning in the matter the developer explained. They highlighted: • That the MRDO value set includes levofloxacin and excludes vancomycin.	The "BroadSpectrumAntibioticsMDRO" value set excludes vancomycin because it refers to any of those antibiotics in oral or intravenous form. Oral vancomycin is a treatment for *C. Difficile* infection, not for MRSA. Thus, we did not want to count this as inappropriate MRSA therapy. Our submitted CQL code describes that IV vancomycin (which is listed as an individual code on our data dictionary attachment) should be considered MDRO therapy in addition to antibiotics from the "BroadSpectrumAntibioticsMDRO" value set. Both ciprofloxacin and levofloxacin are included in the "BroadSpectrumAntibioticsMDRO" value set. Although fluoroquinolones are potential therapy for community-acquired pneumonia, they are associated with antibiotic resistance and severe side effects. Thus, there are safer alternatives available. This is in line

Feedback/Questions	Written Developer Response
	with updated stewardship guidance from IDSA fluoroquinolones be reserved for patients unable to tolerate first line therapy or who have risk factors for MDROs (https://www.idsociety.org/globalassets/idsa/practice-guidelines/community-acquired-pneumonia-in-adults/cap-clinical-pathway-final-online.pdf). We reviewed this question with our TEP and seven out of eight agreed: "We should try to reduce empiric fluoroquinolone use for hospitalized patients with CAP." The TEP was concerned that if we did not include fluoroquinolones in the MDRO antibiotics then people would prescribe them to "game" the metric and get anti-pseudomonal coverage, "I am most concerned about driving up use of FQs, especially levofloxacin because of its guideline recommendation & fact that it is anti-Pseudomonal. I worry that persons concerned about meeting the metric but also worried about Pseudomonas will use levofloxacin more."

CBE #4700e: Full Responses Written by the Developer

Feedback/Questions	Written Developer Response
Follow-up Care: Several Advisory Group members asked for clarification on what would happen if a patient moved to a different health system for a second opinion or follow-up care, particularly if the systems are not using the same electronic medical record (EMR) system. They asked about potential data issues if a patient did not receive care at the same location. Another Advisory Group member stated that it would be helpful to have data on how often patients switch providers between an initial screening and next steps.	We do not have data to quantify the frequency of out-of-system diagnostic resolution for breast cancer screening. However, the rates are low considering the high performance of all three health systems on this eCQM. Our data showed that for patients reaching diagnostic resolution in Health System 1, 89% to 99% received both screening and follow-up diagnostic resolution within the same facility group in 2023, which is the year most reflective of current practice. Stakeholders recommended consideration for exceptions that might be outside of a facility's control or minimizing expectations that performance should always be 100%. In response, we applied a benchmark of 90% to accommodate for out-of-system follow-ups for the performance gap assessment.

Feedback/Questions	Written Developer Response
Timeframe: Several Advisory Group members asked for clarification about what takes place within the 60 days specified in the measure. Several Advisory Group members stated a quick follow-up time helps to alleviate patient anxiety. Others mentioned that rural and under-resourced facilities might require the time allotted in the measure or even longer. One Advisory Group member highlighted that while they personally prefer the 60-day timeframe, this measure does not currently align with a National Committee for Quality Assurance (NCQA) measure, Follow-Up after Abnormal Breast Cancer Assessment (BCF-E).	All steps leading to diagnostic resolution are required to occur within 60 days. Our data from Health Systems 1 and 2 show that it is feasible to complete this process in 60 days with performance rates around the 90% benchmark. We recently received data from Health System 3, which showed an overall rate of 90.5% (95% CI: 89.8%, 91.1%) for one integrated delivery system over 8 years. The CDC's National Breast and Cervical Cancer Early Detection Program (NBCCEDP) currently supports the use of a 60-day timeframe for follow-up completion in the breast cancer screening continuum. Literature has shown that wait times exceeding 90 days to diagnostic follow-up are associated with increased tumor size and lymph node metastases. Furthermore, symptomatic women with delays greater than 90 days have lower survival odds compared to women who are quickly diagnosed and can start treatment earlier.
Value: An Advisory Group member asked what value the measure would create in addition to measures that the NCQA has proposed for the Healthcare Effectiveness Data and Information Set (HEDIS): Documented BI-RADS Assessment after Mammogram (DBM-E) and Follow-Up after Abnormal Breast Cancer Assessment (BCF-E).	This eCQM emphasizes timeliness of diagnostic resolution to detect breast cancers. The measure is patient based. The HEDIS measure targets health plans to report "the percentage of inconclusive or high-risk BI-RADS assessments that received appropriate follow-up within 90 days of the assessment, for members 40–74 years of age." The HEDIS measure is based on mammography episodes, rather than patients. A patient may be included in the denominator multiple times in one measurement year, for example, if the patient received both an inconclusive screening mammogram and high-risk assessment on diagnostic follow-up imaging. The HEDIS measure is likely to face a ceiling effect by allowing for up to 180 days (90 days after inconclusive screening mammogram + 90 days after high-risk assessment on diagnostic follow-up imaging) to diagnostic resolution requiring biopsy. The HEDIS measure includes only diagnostic mammograms and breast ultrasounds as appropriate follow-up after an inconclusive (BI-RADS 0) assessment.

Feedback/Questions	Written Developer Response
	Our eCQM is more comprehensive by also including breast MRI as appropriate diagnostic follow-up.
Cost: A few Advisory Group members expressed concern about how cost may play into the measure and asked what would be covered for a patient beyond the initial mammogram and if there could be misconceptions about the cost of additional testing.	Out-of-pocket costs may affect the likelihood of timely diagnostic resolution of the breast cancer screening process. Several U.S. states have passed legislation mandating insurance providers to cover diagnostic/supplemental imaging [1], including the State of Massachusetts, which recently implemented the Act Relative to Medically Necessary Breast Screenings and Exams for Equity and Early Detection to cover follow-up breast imaging including MRIs and ultrasounds while preventing increases in patient cost-sharing by 2026. This eCQM provides the opportunity to measure differences between groups and potentially influence further policies to cover costs associated with the full breast cancer screening process, including breast biopsy. [1] A State-By-State Look at Diagnostic and Supplemental Breast Imaging. https://www.komen.org/blog/a-state-by-state-look-at-diagnostic-and-supplemental-breast-imaging/
Feasibility: An Advisory Group member asked if all sites have data fields that are discrete enough to capture final diagnosis or if that information is embedded in free text.	EPIC and Cerner have structured fields for last overall assessments, which are the final interpretations of imaging for the reports based on the initial assessment and any addenda, for mammography, breast ultrasound, and breast MRI. Integrated delivery systems/facility groups reporting this eCQM should be encouraged to leverage these fields for straightforward calculation of this measure. As of 2023, all facility groups within Health System 1 used the structured fields for last overall assessment. Otherwise, simple string search or natural language processing (NLP) can be used to mine the impression and addenda of the imaging reports to extract the last overall assessment. This is a straightforward string search/NLP task with near perfect performance.

Feedback/Questions	Written Developer Response
	This approach was applied by Health System 3 to extract data from a legacy system.

CBE #4700e: Full Responses Written by the Developer

Feedback/Questions	Written Developer Response
Patient Reluctancy: A few Advisory Group members stated that this measure may be difficult because patients who opt not to go for a colonoscopy in the first place are often reluctant even after a FIT. They suggested the developer allow exclusions for patients who decline.	We considered this exclusion and obtained stakeholder feedback that the responsibility should be on the providers within the integrated delivery system/facility group to explain the importance of follow-up colonoscopy to patients and encourage timely adherence to this important step of the cancer-screening process. Additionally, excluding patients that declined colonoscopy could place the measure at risk of "gaming" where patients might be considered to have implicitly declined colonoscopy if several outreach attempts to schedule a colonoscopy are not successful.
Consideration of Care Coordination within the Measure: A few Advisory Group members stated that a primary care physician usually (PCP) initiates the stool test, but a specialist would perform the colonoscopy. They asked for more information regarding what happens if the specialist and the PCP are in different systems, how often this occurs, and for the developer to comment on issues related to communications and follow-up bias.	**In specific to out-of-system referrals for colonoscopy:** We do not have data to quantify the frequency of out-of-system referrals for colonoscopy. However, the literature has shown that not accounting for completed out-of-system colonoscopies could underestimate follow-up rates by ~10-13% (relative percentage) [1]. Stakeholders recommended consideration for exceptions that might be outside of a facility's control or minimizing expectations that performance should always be 100%. In response, we applied a benchmark of 80% to accommodate for out-of-system follow-ups for the performance gap assessment. [1] Mohl JT, Ciemins EL, Miller-Wilson LA, Gillen A, Luo R, Colangelo F. Rates of Follow-up Colonoscopy After a Positive Stool-Based Screening Test Result for Colorectal Cancer Among Health Care Organizations in the US, 2017-2020. JAMA Netw Open. 2023;6(1):e2251384. Published 2023 Jan 3. PMID: 36652246.
Cost: An Advisory Group member pointed out that once a patient has a positive stool-based test, the colonoscopy would become diagnostic.	Out-of-pocket costs may affect likelihood of having a missed or delayed follow-up colonoscopy. As of January 1, 2023, Affordable Care

Feedback/Questions	Written Developer Response
They expressed concern that this would be an out-of-pocket cost for many patients. Another Advisory Group member mentioned that even with the new policy changes noted by the developer, they were not sure how consistently ICD-10 codes were being appropriately used on claims, which are needed to avoid the out-of-pocket costs.	Act (ACA)-compliant private insurers and Medicare are required to eliminate cost-sharing for follow-up colonoscopies performed after positive non-invasive stool-based colorectal cancer screening test results. While polyp removal is fully covered by private insurance, Medicare recipients are required to pay 15% of the cost until 2026. This eCQM provides the opportunity to measure differences between groups and potentially influence policy to cover costs associated with the full colorectal cancer screening process. For example, we re-calculated eCQM performance rates stratified by primary insurance and showed significantly higher rates of follow-up colonoscopy in privately insured patients (59.1%; 95% CI: 57.3%, 61.0%) compared with those on Medicare (53.8%; 95% CI: 51.3%, 56.3%) or Medicaid (48.3%; 95% CI: 43.5%, 53.2%). *Please note that these are updated rates. We were able to include previously missing data on insurance for these updated rates. We now have >99% coverage for patient primary insurance at the time of the positive stool-based test.*