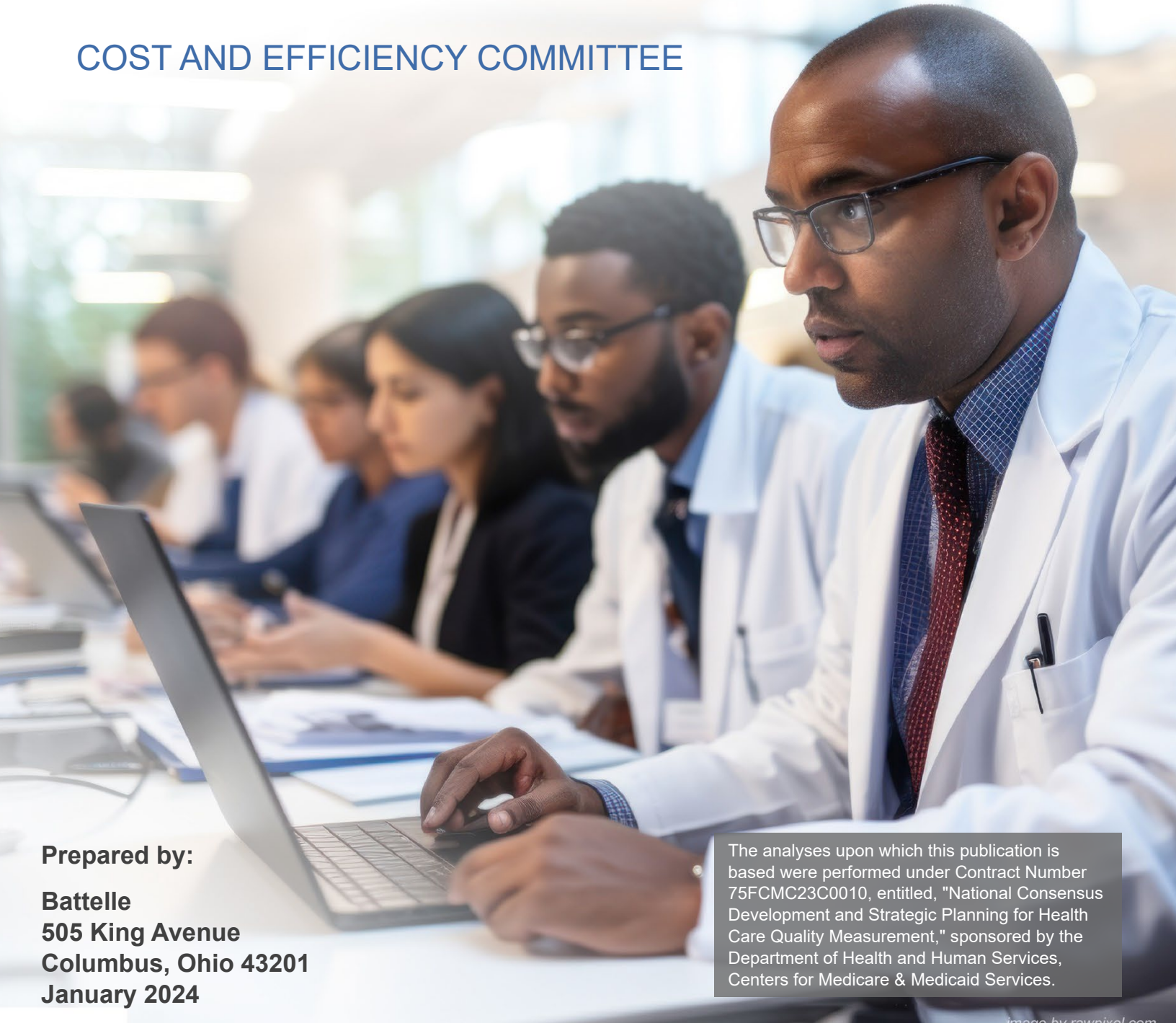
Partnership for Quality Measurement
Powered by Battelle

Fall 2023 Endorsement and Maintenance (E&M) Committee Independent Review Summary

COST AND EFFICIENCY COMMITTEE



Prepared by:

Battelle
505 King Avenue
Columbus, Ohio 43201
January 2024

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

Image by rawpixel.com

Table of Contents

Summary of Committee Independent Reviews

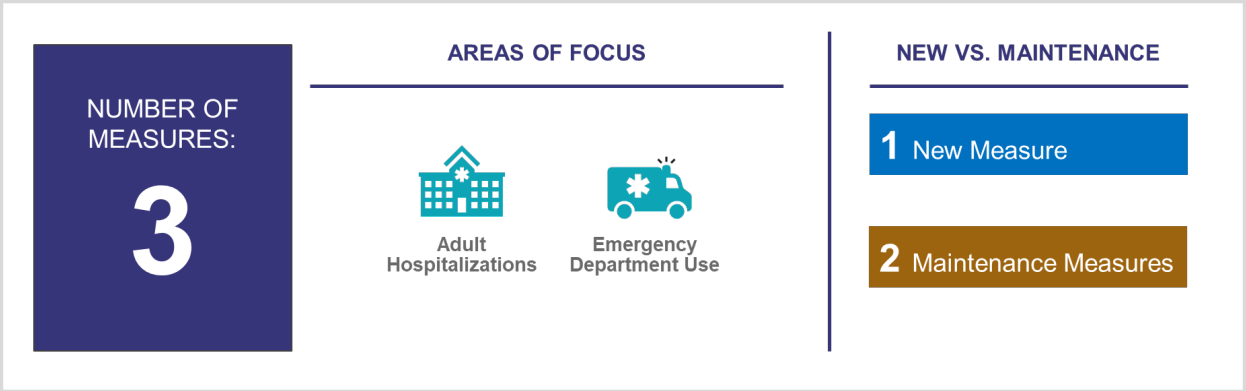
Independent E&M Committee Member Reviews Overview

At least three (3) weeks prior to an E&M committee endorsement meeting, the Recommendations Group and the Advisory Group of each E&M committee receive the full measure submission details for each measure up for review, including all attachments, the Partnership for Quality Measurement (PQM) Measure Evaluation Rubric, the public comments received for the measure(s) under review, and the E&M team preliminary assessments.

Members of both groups were asked to review each measure, independently, against the PQM Measure Evaluation Rubric. Committee members assigned a rating of "Met," "Not Met but Addressable," or "Not Met" for each domain of the PQM Measure Evaluation Rubric. In addition, committee members provided associated rationale for each domain rating, which is based on the rating criteria listed for each domain. Battelle staff aggregated and summarized the results and distributed them back to the committee, and to the respective measure developers and/or stewards, for review within one (1) week of the endorsement meeting.

These independent committee member ratings are compiled and used by Battelle facilitators and committee co-chairs to guide committee discussions.

Figure 1. Fall 2023 Measures for Committee Review



For the Fall 2023 cycle, the Cost and Efficiency committee received three (3) measures, one (1) new measure and two (2) measures undergoing maintenance endorsement review (Figure 1). The measures focused on adult hospitalizations after outpatient surgery or following percutaneous coronary intervention and emergency department use following an inpatient psychiatric facility discharge.

Measure-Specific Summaries

The following brief summaries include themes and considerations gathered from the committee's independent reviews for each of the five domains of the PQM Measure Evaluation Rubric. Themes were assessed and categorized with respect to the strengths and limitations of the measure(s) under endorsement review. Corresponding to the themes are the number of committee reviews received and stratified by the ratings of "Met," "Not Met," and "Not Met, but Addressable."

CBE #0695 – Hospital 30-Day Risk-Standardized Readmission Rates following Percutaneous Coronary Intervention (PCI)

Number of Committee Reviews: 20

Importance (n=20)	Strengths	Limitations
Consensus 10% Met; 10% Not Met, but Addressable; 80% Not Met	• **Importance of Measure:** The measure is important for assessing patient outcomes following Percutaneous Coronary Intervention (PCI) and comparing facilities performing PCIs. Literature, patient feedback, and expert face validity indicate the importance of this measure. • **Association between PCI treatments and complications:** Direct and indirect evidence supports the association between PCI treatments and complications which may lead to readmissions. • **Health System Monitoring:** Tracking readmissions and complications is important for health system monitoring.	• **Limited Literature:** Limited literature is provided justifying the causal relationship between low quality of care (as it relates to Percutaneous Coronary Intervention) and readmissions. • **Current Data:** Data presented is over 12 years old (2010-2011). • **Lack of Patient Involvement:** The measure lacks specific details regarding patient involvement in the measure development process. • **Provider Attribution:** The 30-day readmission window introduces factors into the outcome variable that cannot be solely attributed to provider care. It is suggested that the performing physician or his/her group practice should be the accountable party, not the acute care facility.

Feasibility (n=20)	Strengths	Limitations
No Consensus 50% Met; 25% Not Met, but Addressable; 25% Not Met	• **Data Collection:** Measure is based on electronic claims data and registry information.	• **Registry Participation Barrier:** Acknowledge that some facilities do not/cannot participate in the registry. • **Data Linkage:** Concerns with consistent linking of data from CathPCI Registry and Medicare claims data.

Scientific Acceptability (n=20)	Strengths	Limitations
Consensus 5% Met; 10% Not Met, but Addressable; 85% Not Met	• **Measure Specifications:** There are clearly defined inclusion/exclusion criteria. • **Data Validation:** Data elements from NCDR CathPCI data elements were validated and considered a reliable clinical source. Approach to validity testing using the ACC's audit program appears reasonable to confirm the accuracy of the data elements in the registry.	• **Outdated Data:** Data presented is over 12 years old (2010-2011). • **Concerns with Numerator and Denominator:** The denominator doesn't include PCIs performed in outpatient clinics and excludes facilities with less than 25 procedures. • No clinical or statistical justification as to why the 30-day window was used. Only hospital readmissions are included. If a patient dies after discharge from the hospital, that is also treated as a success since there is no readmission. • The denominator only includes PCIs performed in hospitals, not PCIs performed in ambulatory surgery centers (ASCs).

Scientific Acceptability (n=20)	Strengths	Limitations
		• **Reliability:** Reliability testing falls below the acceptable 0.6 threshold (i.e., split-half reliability ICC of 0.3711). • **Validity:** The validity and discriminatory statistics are poor. Validity is asserted but not proven/identified. The submission lacks the necessary data to ensure its scientific validity.

Equity (n=20)	Strengths	Limitations
No Consensus 5% Met; 30% Not Met, but Addressable; 65% Not Met	• None	• **Lack of Equity Information:** No information provided by developer. Concerns with equity implications and lack of discussion of social determinant of health. Recommendations for an analysis of demographic data included in the registry. • **Percutaneous Coronary Intervention (PCI):** Comments mention that conditions requiring PCI are much more prevalent in populations of color and lower socioeconomic patients. A measure not taking these factors into account for this treatment is not useful for many patients most in need of the information.

Use and Usability (n=20)	Strengths	Limitations
Consensus 5% Met; 5% Not Met, but Addressable; 90% Not Met	• **Support to Participants:** Support is provided to participants including calls, conferences, and support from clinical quality associates.	• **Not Currently in Use:** Measure is not in use. • **Limited Use:** Concerns with limited application to a subset of facilities (i.e., those that perform PCIs and participate in the CathPCI registry, no ambulatory surgery centers) and readmission measurement for a subset of patients (i.e., over 65 on Medicare). • **Data Issues:** The data set is outdated, with no current data indicating a clinical concern that needs to be addressed. There are also problems in connecting two data sources that have not been connected for a decade. • **Feedback and Improvement:** There is no information on feedback on measures, considerations from measure feedback, progress on improvement, and unexpected findings.

CBE #4190 – 30-Day Risk Standardized All-Cause Emergency Department Visit Following an Inpatient Psychiatric Facility Discharge

Number of Committee Reviews: 21

Importance (n=21)	Strengths	Limitations
No Consensus 43% Met; 29% Not Met, but Addressable; 29% Not Met	• **Emergency Department (ED) Visits and Readmissions:** There's a significant correlation between ED visits and readmissions in IPFs, suggesting a potential to merge these measures. • **Follow-up After Discharge:** Post-discharge follow-ups are crucial, with a community initiative increasing 30-day follow-up rates from 18% to 75%. • **Evidence Base and Relevance to the Measure:** Most literature supports reducing readmissions over just ED visits post-discharge in IPFs. This measure is backed by research and expert panels. • **Inclusion of Behavioral Health (BH) Conditions and Patients:** Including BH conditions and readmissions provides a comprehensive view of a BH patient's post-discharge experience. Long-term stability is a key goal for high-risk BH patients. • **Program IPF's Influence on Outcomes:** The developer has effectively outlined this measure's importance. It's exciting to review care quality and	• **Emergency Department (ED) Visits and Readmissions:** The measure includes all ED visits by adults, potentially skewing results. Unrelated visits without behavioral health issues aren't counted. • The timeframe for visits (30-days) isn't clear. The assumption of a 5% reduction in ED visits lacks justification, making the impact assessment on healthcare costs invalid. Cited studies are weak observational studies. • **Follow-up After Discharge:** Little evidence is provided on the significance of ED returns as a care gap. A measure on non-follow-up visits scheduled before discharge could be more clinically relevant. • **Program IPF's Influence on Outcomes:** The measure is useful but limited to program IPF, which may not influence outcomes. The TEP included technical experts and patient caregivers, but the information provided, especially from the dissenting caregiver, was incomplete. The measure's meaningfulness remains unclear.

Importance (n=21)	Strengths	Limitations
	transitions from IPFs. More input is needed, but adding ER admission tracking is vital for patients.	

Feasibility (n=21)	Strengths	Limitations
Consensus 90% Met; 0% Not Met, but Addressable; 10% Not Met	• **Claims-Based Measure:** The measure is a claims-based measure and is feasible to collect. The system to collect data and calculate the measure is an automated process using electronic standardized data already routinely generated for billing purposes. • **Data Availability:** Data are readily available. The measure relies on readily available Medicare claims data, and no data availability issues were identified.	• **Feasibility Concerns:** The measure does not meet feasibility criteria as long-term or no path is specified to support routine and electronic data capture with an implementable data collection strategy. Data collection can be onerous on already resource-challenged facilities. • **Patient Population and ED Visits:** The patient population included in this metric, particularly those who live in inner city areas, frequently visit multiple EDs belonging to different health systems, which reduces the ability of any one facility to reduce these ED visits significantly.

Scientific Acceptability (n=21)	Strengths	Limitations
No Consensus 14% Met;	• **Specifications Well-Defined:** Measure is well defined and specified.	• **Reliability Testing:** Nearly 50% of the entities could have a reliability <0.6.

Scientific Acceptability (n=21)	Strengths	Limitations
48% Not Met, but Addressable; 38% Not Met	• **Reliability Testing:** Reliability testing shows over 50% of entities (from a 2019-2021 dataset of 1,483 entities) have a reliability >0.6. • **Measure Validity:** The measure's validity is supported by 12 published references, showing its association with discharge planning and post-discharge care. • **Risk Adjustment:** The risk adjustment approach is robust, with a C-statistic of 0.67, similar to other models, and uses a strong method for calibration.	• **Validity Testing:** For face validity testing, which is acceptable for new measures, the submission references a TEP (N=7) but does not report results. • The submission references a TEP for face validity testing but doesn't report results. Empirical validity testing was performed with modest effects, but no rationale explaining the results based on the hypothesis was provided. • **Concerns with Numerator:** Concerns about all cause ED use rather than mental health associated visits. • **Risk Adjustment:** The risk adjustment model has serious flaws, including inadequate performance evaluation and failure to adjust for availability of community mental health care; for whether the patient was discharged to another facility; for propensity of patients to use the ED for chronic conditions; for factors affecting patient access to other types of health services.

Equity (n=21)	Strengths	Limitations
No Consensus 62% Met;	• **Social Determinants of Health (SDOH):** The developer evaluated 17 SDOH for disparities in the measure. The measure developers gathered data on SDOH and risk-stratified data by SDOH. Multiple SDOH are tracked, enabling more detailed analysis.	• **Social Determinants of Health (SDOH):** SDOH variables were added to the risk model but had weak associations with the outcome and were not retained. • **Equity:** The measure doesn't evaluate performance related to reducing health care inequities, and it doesn't focus on

Equity (n=21)	Strengths	Limitations
24% Not Met, but Addressable; 14% Not Met	• **Equity:** Some respondents thought the assessment for equity was sufficient and that the developer did a good job in portraying the contribution to health equity.	healthcare disparities. Readmissions and outcomes could vary widely based on social factors. • **Access to Intensive Psychiatric Facility (IPF)**: The study did not sufficiently address access to IPF, which could be directly correlated to BH ED admissions.

Use and Usability (n=21)	Strengths	Limitations
No Consensus 14% Met; 48% Not Met, but Addressable; 38% Not Met	• **Plan for Use:** The measure is planned for use in public reporting and internal/external QI. • **Strategies for Improving Post-Discharge Continuity of Care:** Developer suggests that IPFs focus on implementing strategies for improving post-discharge continuity of care.	• **Measure Utility:** Concerns about actionable information for improvement. Suggestion to include ED visits that result in admissions. • **Quality Improvement vs. Accountability:** Measure seen as usable for quality improvement but not for provider comparison. • **Measure Adjustments:** Suggestion that the measure's usability would improve with some adjustments, such as limiting to BH diagnoses and handling AMA-discharges. There's also a call for adjustments to the inclusion/exclusion criteria, such as not excluding patients who died after discharge.

CBE #2687 – Hospital Visits after Outpatient Surgery

Number of Committee Reviews: 21

Importance (n=21)	Strengths	Limitations
No Consensus 67% Met; 19% Not Met, but Addressable; 14% Not Met	• **Importance of the Measure:** The measure is crucial for hospitals and patients, driving quality improvements by reducing adverse outcomes related to same-day surgery preparation. • **Evidence Supporting the Measure:** Developers provided evidence supporting the measure, highlighting variability across departments and potential interventions. • **Benefit to the Measure:** 12 of the 13 TEP members moderately or strongly agreed that the measure can be used to distinguish between better and worse quality. • **Significance of the Measure**: The measure is significant as it provides insights into outpatient surgery care quality, examining potential post-surgery issues.	• **Lack of Specificity:** The current measure lacks specificity in areas such as non-hospital mortalities and urgent care visits. • Additional specifications for ED visits within 7 days could strengthen the measure and prevent penalizing overcautious patients and physicians. • The measure doesn't calculate quality separately for specific surgeries, limiting improvement insights. • All-cause events should be limited to the first 72 hours post discharge. This is a serious design flaw and injures the face validity of the measure. Not sure a global measure for all surgeries make sense - a facility with lots of cataracts will look different from a safety net hospital. • **Lack of Patient/Caregiver Input:** No information about the number of patients and caregivers in TEP or what their comments were.

Feasibility (n=21)	Strengths	Limitations
Consensus 95% Met;	• **Claims-Based Measure:** The measure is based on claims data. This approach is feasible, reduces provider burden, and has been in use for several years.	• **Feasibility Concerns:** All-cause events should be limited to the first 72 hours post discharge.

Feasibility (n=21)	Strengths	Limitations
0% Not Met, but Addressable; 5% Not Met	• **Automated Process:** The system to collect data and calculate the measure is an automated process using electronic standardized data. All data required as specified in this proposal is in the EMR. • **No Additional Burden or Fees:** The developer mentioned there are no fees, licensing, or other requirements to use this measure as specified.	

Scientific Acceptability (n=21)	Strengths	Limitations
No Consensus 57% Met; 14% Not Met, but Addressable; 29% Not Met	• **Reliability Testing:** Facility-level reliability tested is completed with recent data (2022). Results indicate that the majority of facilities scored above the 0.6 threshold. • **Validity Testing - Correlation with Surgical Volume:** The developer presented the correlation between the measure and a related performance measure (surgical volume). There was an overall trend toward improved outcomes with increasing volume. The correlation coefficient between facility-level procedural volume and the hospital outpatient surgery measure score was -0.18, as hypothesized.	• **Numerator and Denominator:** The measure's broad denominator and numerator that includes unrelated events can skew post-surgery visit rates. Concern with "all cause" readmission being used rather than more specific readmission rates. • **Empirical Validity:** The empirical validity for this measure does not support the convergent validity of the measure. The correlation between this measure and the criterion measure is very weak and not statistically significant (0.033; p=0.07). • **Model Comparison:** There is mention of the OP-32 colonoscopy measure but no additional comparison or mention if there is duplicity between this all-procedure measure vs. OP-32. Would recommend comparing model performance for the aggregated all procedure vs. individual

Scientific Acceptability (n=21)	Strengths	Limitations
	• The TEP face validity check is okay. • **Risk Adjustment:** The risk adjustment seems sufficient.	groups of same/similar procedures to ensure complete information is captured and variation explained by the model. • **Risk Adjustment:** Comments note concerns with modeling approach and that several variables not addressed in the model (e.g., failure to adjust for outpatient surgeries performed at other facilities, for the propensity of patients to visit the hospital for other health problems, for type of surgery). A single model may be insufficient. • The model uses surgical site but not surgical intensity or risk, raising concerns about noise and predictive power. • Concerns are raised about hospital classification due to over- and under-prediction. • The discriminatory c-statistic of 0.693 is not sufficient for a measure tied to payment.

Equity (n=21)	Strengths	Limitations
No Consensus 71% Met; 14% Not Met, but Addressable;	• **Dual Eligibility (DE) and Area Deprivation Index (ADI):** The developer's risk adjusted results based on dual eligibility status and area deprivation index are adequate.	• **Support Further Analysis:** Comments support further analysis related to social determinants reported due to likely effect on the measure outcome and patient care. • **Problematic Hospital Classification System:** Comments note measure methodology could inappropriately classify hospitals as “worse than expected.” Relatedly, treatment of people within marginalized groups (those without access to

Equity (n=21)	Strengths	Limitations
14% Not Met	• Correlation analysis indicates that adjusting for area-based socioeconomic measures would have minimal impact on rankings. • **Stratification of the Measure:** Developers are addressing health care disparities by implementing a stratification methodology. • The measure is stratified using within- and between-hospital comparisons. • Stratifying by dual eligibility adequately addresses equity issues related to socioeconomic status variance.	primary care) could be affected by measure classification system.

Use and Usability (n=21)	Strengths	Limitations
No Consensus 62% Met; 19% Not Met, but Addressable; 19% Not Met	• **Measure Currently in Use:** The measure is currently in use in the HOQR. • **Feedback Mechanism:** Feedback on the measure can be submitted via Quality Net. Facilities receive confidential, detailed reports outlining patient-level information (e.g., unplanned visits, performance relative to state and national benchmarks).	• **Patient Impact:** It is not clear that all factors affecting patient satisfaction are addressed. A patient who is choosing where to have surgery wants to know whether a hospital delivers high-quality care for that specific type of surgery. This could potentially lead patients to avoid needed surgery or choose ill-equipped facilities. • **Incentive Issues:** The measure's weaknesses could lead to hospitals being mislabeled as "worse than expected," creating

Use and Usability (n=21)	Strengths	Limitations
		an incentive to avoid treating patients likely to have higher visit numbers.


Partnership for
Quality Measurement
Powered by Battelle