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

MANAGEMENT OF ACUTE EVENTS AND CHRONIC CONDITIONS COMMITTEE

February 7, 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

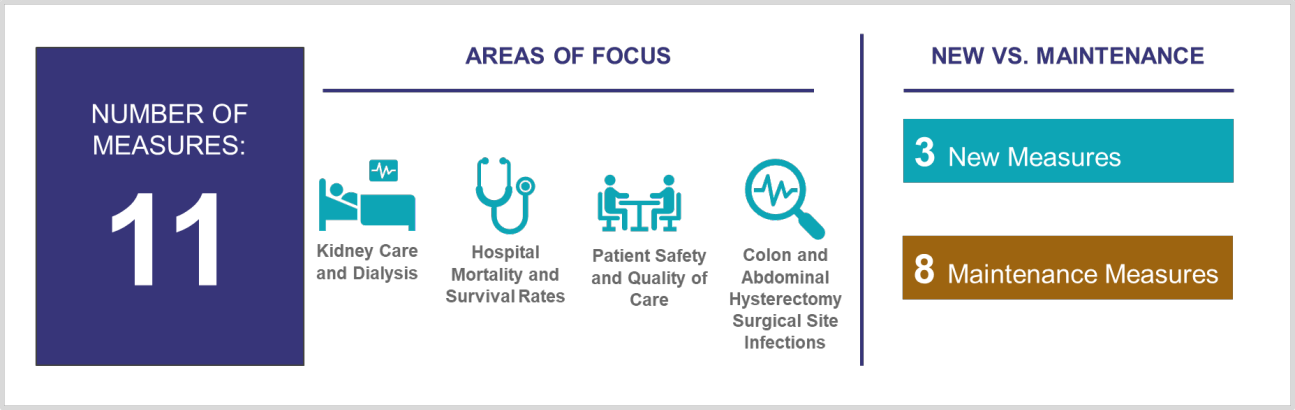
Overview of Fall 2024 Measures for Review

During this measure review cycle, developers and stewards submitted 11 measures to the Management of Acute Events and Chronic Conditions committee for endorsement consideration (Table 1). The measures focused on kidney care and dialysis, hospital mortality and survival rates, patient safety and quality of care, and colon and abdominal hysterectomy surgical site infections. (Figure 1).

Table 1. Overview of Measures Under Endorsement Review

CBE Number	Measure Title	New/Maintenance	Developer/Steward
0318	Delivered Dose of Peritoneal Dialysis Above Minimum	Maintenance	University of Michigan (UMICH)/Centers for Medicare & Medicaid Services (CMS)
0531	Patient Safety Indicator (PSI) 90: Patient Safety and Adverse Events Composite	Maintenance	Mathematica/CMS
0753	30-Day Post-Operative Colon Surgery (COLO) and Abdominal Hysterectomy (HYST) Surgical Site Infection (SSI) Standardized Infection Ration (SIR)	Maintenance	Centers for Disease Control and Prevention (CDC)
1423	Minimum spKt/V for Pediatric Hemodialysis Patients	Maintenance	UMICH/CMS
1425	Measurement of nPCR for Pediatric Hemodialysis Patients	Maintenance	UMICH/CMS
2706	Pediatric Peritoneal Dialysis Adequacy: Achievement of Target Kt/V	Maintenance	UMICH/CMS
3309	Risk-Standardized Survival Rate (RSSR) for In-Hospital Cardiac Arrest	Maintenance	American Heart Association (AHA)
3502e	Hybrid Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure with Claims and Electronic Health Record Data	Maintenance	Yale Center for Outcomes Research and Evaluation (Yale CORE)/CMS
4580	Composite measure for the quality of care provided to patients undergoing percutaneous coronary interventions (PCI)	New	American College of Cardiology (ACC)
4595	Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Acute Ischemic Stroke Hospitalization with Claims-Based Risk Adjustment for Stroke Severity	New	Yale CORE/CMS
4650	Prevention of Chronic Hyperphosphatemia in Dialysis Patients	New	UMICH/CMS

Figure 1. Fall 2024 Measures for Committee Review

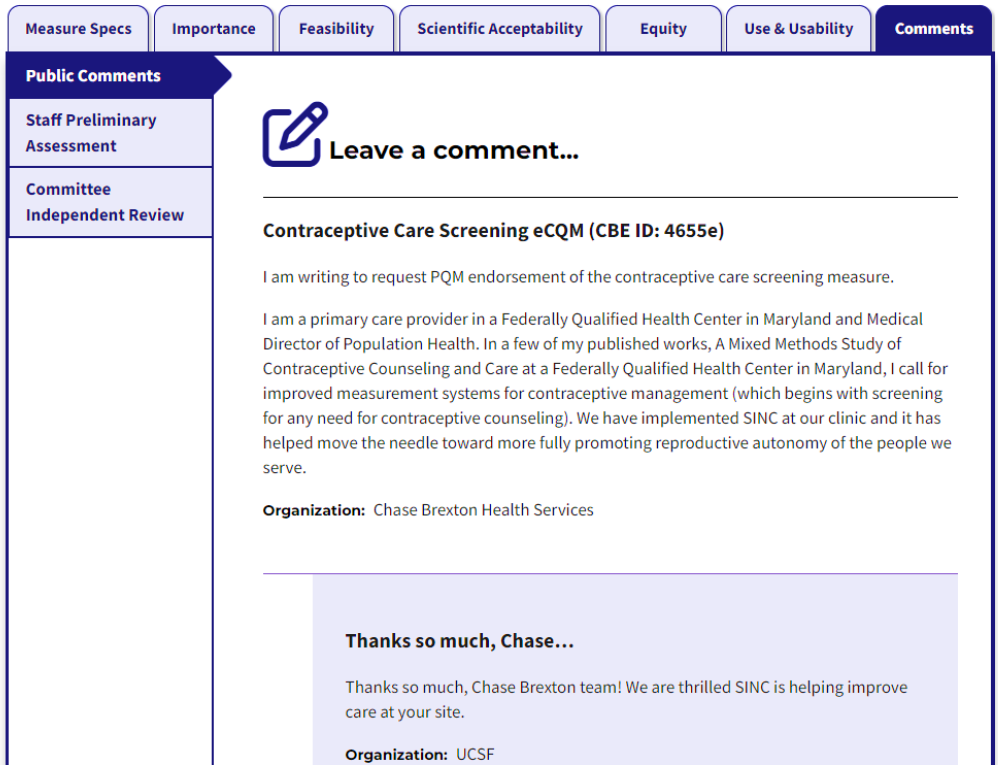


Public Comment

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 2, 2024. Thirty of 35 (86%) active Advisory Group members attended to share feedback and ask questions regarding the measures under endorsement review. Developers of the respective measures also attended and provided responses to the Advisory Group questions. After the meeting, developers had the opportunity to submit additional written responses to Advisory Group member feedback and questions (Appendix A).

The measure evaluation summaries in this discussion guide contain overviews of the Advisory Group member discussions and developer 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 #0318: Delivered Dose of Peritoneal Dialysis Above Minimum [UMICH/CMS]

Specifications

Measure Description: *Percentage of all patient months for adult patients (>= 18 years old) whose delivered peritoneal dialysis dose was a weekly Kt/V urea >= 1.7 (dialytic + residual).*

Staff Preliminary Assessment Rating[1]

Importance: Met

Rationale: This maintenance measure meets all criteria for "Met" due to its robust, well-graded evidence base, clear business case, documented performance gap, and meaningfulness to patients, making it essential for addressing peritoneal dialysis among adults.

Feasibility: Met

Rationale: This maintenance 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 health care system.

Reliability: Met

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

Validity: Not Met but Addressable

Rationale: The developer provides empirical evidence of a facility-level co-variation between the measure focus and two material outcomes, mortality (not statistically significant) and hospitalization (statistically significant) in the causal chain. Going forward, additional studies that either rule out potential confounding factors or describe features of potential mechanisms will strengthen causal claims.

Equity: Not Met

Rationale: The developer did not address this optional domain.

Use & Usability: Met

Rationale: For maintenance, the measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. From the initial implementation, the measure demonstrated an increase in performance results, affirming its ongoing usability. The developer reports no unexpected findings.

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

Public Comment

This measure did not receive any comments during the public comment period.

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Age of Guidelines: An Advisory Group member pointed out that the guidelines the measure is based upon are from 2006. They asked if the field has more current guidelines.	The evidence has not changed since then.
Scientific Acceptability: An Advisory Group member pointed out that patients are counted multiple times if they switch providers. They asked for clarification on how that affects the measure.	Patients are counted monthly, meaning that patients can contribute to the measure 12 times a year. (They noted that efficacy for peritoneal dialysis is typically done every 3 months.) If a patient moves to a different facility, the count starts over so that one facility is not held accountable for the actions of another.
Integration into the End-Stage Renal Disease (ESRD) Quality Improvement Program (QIP): An Advisory Group member asked for more information on how the measure is scored and integrated into its QIP.	Battelle staff noted that such questions may be more appropriate for the CMS program team.
Shared Decision-Making and the 1.7 Metric Goal: A few Advisory Group members shared that some patients may appear to do well on dialysis even without hitting the 1.7 threshold mentioned in the measure. The committee discussed whether these patients should have the goal "imposed" upon them and the difficulty of incorporating "subjective" considerations into a metric. The committee, particularly the patient participants, expressed appreciation for shared decision-making and emphasized that the earlier the patient is brought into the discussion, the better their outcomes will be. Another Advisory Group member said it is important for clinicians to help their patients understand the evidence surrounding dialysis.	Patients value having a voice in their care. The developer is carefully considering how to handle this issue in current and future measure development. Patient advocates are included on the TEP, and their input is considered alongside other feedback collected through public comment.
Compliance: An Advisory Group member asked for the percentage of compliance with the measure.	Because the measure is in QIP, the adherence is fairly high, with the vast majority of patients and providers completing the assessments. About 30% of facilities have achievement rates of less than 90%.
Topped Out: A few Advisory Group members asked, given that the measure has been around for so long, whether it is making a substantial impact or perhaps has become topped out? A few Advisory Group members expressed that they believed it was time to focus on	As a safety measure, a small but meaningful percentage fall below the target helps create a safety net to make sure that those individuals are receiving adequate treatment. They have seen some small improvements over time, acknowledging that the change is not rapid.

Feedback/Questions	Summary of Developer Response
other potential proxies to address this potential gap in care, considering the age of the measure and its high adherence rate.	
Measure Parsimony: An Advisory Group member asked about other potential measures and how the developer considers parsimony.	They did not believe any other measures examined the same aspects of care, pointing out that some other measures may touch upon different types of dialysis or different populations.
Risk Adjustment: An Advisory Group member expressed concern about the lack of risk adjustment in the measures stewarded by the University of Michigan. They asked whether the developer had done an analysis on whether the scores would be misclassified for under-resourced communities. Several Advisory Group members echoed concerns about risk adjustment.	Developers risk adjust measures when evidence suggests it would affect the outcomes or when a technical expert panel (TEP) advises risk adjustment. In the case of the dialysis measures, patient-level or regional factors are unlikely to affect the measure, as the dialysis facility provides supplies and prescriptions. Further, the developer has not been advised to risk adjust the measure.
Validity: An Advisory Group member asked the developer to comment on the validity for this measure not being statistically significant.	The main issue they have encountered for validity is that most facilities provide hemodialysis and peritoneal dialysis. They had a difficult time demonstrating differences in mortality or hospitalization because hemodialysis overshadows peritoneal dialysis. They had statistical significance for hospitalization.
Reliability: An Advisory Group member commented that they had a difficult time following the reliability methodology. They asked for clarification on whether the minimum is truly identical in 126 facilities.	The inter-unit reliability (IUR) was fairly strong and perhaps there was a presentation issue. It is correct that a large group did not have the same minimum IUR.
How Facilities Can Improve: An Advisory Group member asked the developer how facilities can help patients improve their targeted dialysis number?	There are multiple ways to change that prescription that can be discussed with the patient to suit the patient's needs.
Health Literacy: A patient Advisory Group member commented that this measure may be difficult to understand for individuals who do not have a high degree of knowledge related to dialysis or health literacy.	The developer did not respond to this comment.
Composite Measure: An Advisory Group member said that measures CBE #0318, #4650, #1423, #1425, and #2706 appeared to have a similar focus and asked whether there would be value in considering a composite measure.	The developer did not respond to this comment.

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

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Dissenting	Flexibility for Shared-Decision Making	Advisory Group	The Advisory Group questioned whether the measure left enough flexibility for patient-provider shared decision-making, as some patients appeared to be doing well on dialysis without having hit the 1.7 threshold.

Discussion Category	Key Themes	Source of Comment	Summary of Comments
	Risk Adjustment and Validity Testing	Advisory Group; Staff Assessment	The Advisory Group questioned whether the measure should be risk adjusted. While a risk adjustment evaluation is not required for intermediate outcome measures, a discussion of risk adjustment would strengthen the submission. The developer provides empirical evidence of a facility-level co-variation between the measure focus and two material outcomes, mortality (not statistically significant) and hospitalization (statistically significant). Additional studies are recommended that either rule-out potential confounding factors or describe features of potential mechanisms that will strengthen causal claims.
	Performance Gap	Advisory Group	Given the measure's age and high adherence rate, some Advisory Group members questioned whether the measure is still making a difference or whether it would be appropriate to shift focus to close any potentially still-existing performance gap.

CBE #4650: Prevention of Chronic Hyperphosphatemia in Dialysis Patients [UMICH/CMS]

Specifications

Measure Description: *Percentage of adult dialysis patients with a 6-month rolling average phosphorus value greater than or equal to 6.5 mg/dL.*

Staff Preliminary Assessment Rating

Importance: Met

Rationale: This new measure meets all criteria for "Met" due to its robust, well-graded evidence base, clear business case, documented performance gap, significant anticipated impact, well-articulated logic model, and it addresses the lack of existing measures, making it essential for addressing chronic hyperphosphatemia in dialysis patients.

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 (none specified), ensuring practical implementation within the health care system.

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 moderate 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. The validity claim would be further strengthened by also ruling out other potential causes of co-variation with mortality and hospitalization. The claim can be further supported by confirming a mechanism responsible for the increase seen in outcomes (mortality than hospitalization).

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: 2

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
Gaming: An Advisory Group member asked how stable the patient populations are at dialysis centers, or if a turnover rate could contribute to potential gaming of the measure?	Across the measure's 6-month timeframe, 90% of patients are stable and at the same facility, resulting in little dropout. They felt confident the measure captures the majority of patients.
Hypophosphatemia: An Advisory Group member said they believed that hypophosphatemia should also be measured and addressed.	The TEP was concerned about overcontrolling for phosphorus and that few patients on dialysis have very low phosphorus values. They added that those who do often have low protein or are underweight with a low body mass index (BMI); in accordance, the measure tries to account for patients who are nutritionally impaired by excluding those with low BMI or low serum albumin.
Impact of Social Determinants and Resources on Phosphorus Levels: A few Advisory Group members expressed concern that some facilities would have limited ability to impact phosphorus levels. They pointed out that high phosphorus levels often are a result of having limited access to food, resulting in patients eating more processed food, which is higher in phosphorus. They added that not all facilities will have the same nutrition and dietary resources to support patients.	Facilities can undertake numerous interventions that are supported by the literature. Nutritional counseling is the backbone of helping patients with phosphorus levels and patients can be prescribed phosphorus binders. They looked at the area deprivation index and how much variation there is in phosphorus levels based on a community's access to healthy foods and they found it is not a significant driver in phosphorus outcomes.
Exclusions: An Advisory Group member asked about exclusions for patients with other illnesses that might affect phosphorus levels.	BMI and serum albumin were the two their TEP felt most strongly about.
Care Setting: An Advisory Group member asked for clarification on the definition of "Other Care Setting."	The care setting is the dialysis setting. The response was likely a limitation of the online form.
Potential Burden: An Advisory Group member asked about the responsibility for collecting the information needed for the measure.	This measure (and all the University of Michigan measures discussed at this meeting) rely on data that facilities are already required to report to CMS, which minimizes any additional burden on those facilities.
Risk Adjustment: An Advisory Group member continued the discussion surrounding risk adjustment that began with CBE #0318. They said they believe this measure should also be risk adjusted and that patient-level characteristics likely come into play.	They looked at risk adjustment multiple ways. They examined patient-level factors, including demographics and comorbidities, as well as facility-level factors. However, they found that some risk factors went in the opposite direction of what was anticipated and that they did not find connections they felt would justify risk adjustment. They also

Feedback/Questions	Summary of Developer Response
	discussed this with their TEP. As more data become available, they could consider whether food insecurity would be appropriate to add to the risk adjustment.
Variation Across Facilities: An Advisory Group member asked how performance varies across different facilities, adding that if performance is poor, that means this is likely a good measure on which to focus.	They found huge variation across providers and their ability to control phosphorus levels, resulting in a large performance gap.
January 1 Changes to Phosphorus-Controlling Agents: Several Advisory Group members commented on how, beginning January 1, 2025, dialysis facilities will be responsible for the cost of phosphate binders. Some asked how this may affect patients' access to the binders and how the change will interact with the measure.	Many patients struggle to afford binders under the current system; therefore, a segment of patients will benefit from the change. However, as facilities will now be responsible for cost, they may also be less likely to prescribe more expensive medications. They are not sure how this will play out yet, but that there is a complex interplay between access and choice.
Phosphorus Value of 6.5: A few Advisory Group members wanted more information about how the value of 6.5 had been selected.	They selected 6.5 based on literature, which suggests that values higher than 6.5 are associated with higher risks, including cardiovascular disease issues. They performed sensitivity analyses that indicated even if the measure uses a higher value, a similar patient population is identified.
Importance to Patients: An Advisory Group member commented that they did not feel the evidence provided showed this measure would matter to patients.	High phosphorus levels are associated with cardiovascular issues, morbidity, mortality, and quality-of-life issues such as itching, which is a symptom many dialysis patients deal with.
Measure's Use in the End-Stage Renal Disease (ESRD) Quality Improvement Program (QIP): An Advisory Group member asked for more information on how the measure is scored and integrated into its ESRD QIP.	Battelle leadership noted that such questions may be more appropriate for CMS, and that Battelle convenes Pre-Rulemaking Measure Review committees to consider program use evaluations.
Reliability Table: An Advisory Group member asked for clarification on the results in Table 2, which seem to show that the smallest facilities do the best job, and the largest facilities do the worst. This finding goes against a lot of evidence from other settings that more experienced, higher-volume facilities do better on most quality measures.	After additional review, they discovered that the performance scores in Table 2 were inadvertently copied from Table 1. A corrected version of Table 2 appears in Appendix A.±
Health Literacy: A patient Advisory Group member commented that this measure may be difficult to understand for individuals who do not have a high degree of knowledge related to dialysis or health literacy.	The developer did not respond to this comment.
Composite Measure: Please see description of topic under #0318.	Please see response under #0318.

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

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

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Dissenting	Risk Adjustment and Validity Testing	Advisory Group; Staff Assessment	The Advisory Group questioned whether the measure should be risk adjusted. While a risk adjustment evaluation is not required for intermediate outcome measures, a discussion of risk adjustment would strengthen the submission. The developer provides empirical evidence of a facility-level co-variation between the measure focus and two material outcomes, mortality and hospitalization. Additional studies are recommended that either rule-out potential confounding factors or describe features of potential mechanisms that will strengthen causal claims.
	Facility's Ability to Impact Performance	Advisory Group	The Advisory Group expressed concern that phosphorus levels are affected more by food insecurity and other community-based issues and that facilities would have limited means to change them.
Mixed	Evidence	Public Comment	A public comment from the American Society of Nephrology expressed concerns about the lack of evidence supporting the proposed metric, as there are no clinical trials establishing a specific serum phosphate level target to improve patient outcomes, and the hypoalbuminemia exception is not backed by quality data. However, Kidney Care Partners acknowledges the importance of addressing bone and mineral disorders in kidney failure patients, despite disagreements on the appropriate phosphorus target level, due to the risks associated with unmanaged hyperphosphatemia.
Probing	Upcoming Changes	Advisory Group	The Advisory Group discussed that January 1 changes to how phosphate binders are covered might make this a more important measure to track.

CBE #1423: Minimum spKt/V[2] for Pediatric Hemodialysis Patients [UMICH/CMS]

Specifications

Measure Description: *Percentage of patient months for all pediatric (<18 years old) in-center hemodialysis patients in which the delivered dose of hemodialysis (calculated from the last measurement of the month using the Urea Kinetic Modeling [UKM] or Daugirdas II formula) was spKt/V >= 1.2.*

Staff Preliminary Assessment Rating

Importance: Met

Rationale: This maintenance measure meets all criteria for "Met" due to its robust, well-graded evidence base, clear business case, documented performance gap, and meaningfulness to patients, making it essential for addressing hemodialysis in pediatric populations.

Feasibility: Met

Rationale: This maintenance 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 health care system.

Reliability: Met

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

Validity: Not Met

Rationale: The developer conducted person- or episode-level validity testing on key data elements from 300 facilities, showing high agreement and low missing values, but did not perform accountable entity-level validity testing due to small pediatric sample sizes and did not systematically assess face validity. The absence of material variation of an intermediate outcome measure among entities in the Importance Table (Table 1) also does not support a validity claim. Going forward, a more robust logic model substantiated with face validity from the TEP would provide some modest support for a validity claim.

Equity: Not Met

Rationale: The developer did not address this optional domain.

Use & Usability: Met

Rationale: For maintenance, the measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. From the initial implementation, the measure demonstrated an increase in performance results, affirming its ongoing usability. This was followed by a decline in performance; however, the developer noted that the number of reporting facilities is very small, which can impact stability of performance scores over time. Despite this fluctuation, the developer reports no unexpected findings.

[2] Single-pool urea Kt/V (spKt/V) measures how effectively a single dialysis session removes small waste molecules from your blood.

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 of Guidelines: An Advisory Group member noted that the guidelines are dated and asked if more updated guidelines are available.	The guidelines have not been updated and no new studies have occurred. Because the pediatric population is small, high-quality studies are few.
Importance: A few Advisory Group members noted that this measure is particularly important, because many people are unaware that ESRD affects pediatric patients differently. While the committee acknowledged that the evidence may not be as strong for this measure, this measure affects some of the sickest and most vulnerable patients.	Not applicable.
Hospital Setting: An Advisory Group member noted that pediatric patients often have dialysis in hospital-based settings and wanted to ensure the measure includes those settings.	The measure includes hospital-based settings.
Health Literacy: A patient participant Advisory Group member commented that this measure may be difficult to understand for individuals who do not have a high degree of knowledge related to dialysis or health literacy.	The developer did not respond to this comment.
Risk Adjustment: An Advisory Group member expressed concern about the lack of risk adjustment in the measures stewarded by the University of Michigan.	Please see response under #0318.
Composite Measure: Please see description of topic under #0318.	Please see response under #0318.

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

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance	Advisory Group; Public Comment	The Advisory Group highlighted that this measure is important for the pediatric population and discussed that while the evidence might not be as robust because of the limited population size, these patients are some of the most vulnerable.

Discussion Category	Key Themes	Source of Comment	Summary of Comments
			The American Society of Pediatric Nephrology, American Society of Nephrology, and Kidney Care Partners all support the continued endorsement of pediatric dialysis adequacy measures CBE #1423 and CBE #2706, emphasizing their alignment with KDOQI guidelines and their importance for accountability in treating pediatric patients, while recommending maintaining the pediatric-specific Kt/V goal rather than aligning with adult standards, despite the measures potentially being topped out.
Dissenting	Validity Testing and Risk Adjustment	Advisory Group; Staff Assessment	The Advisory Group questioned whether the measure should be risk adjusted. While a risk adjustment evaluation is not required for intermediate outcome measures, a discussion of risk adjustment would strengthen the submission. The developer conducted person- or episode-level validity testing on key data elements from 300 facilities, showing high agreement and low missing values, but did not perform accountable-entity level validity testing due to small pediatric sample sizes, which is evidence with narrow variation seen in performance gap Table 1. The developer also did not systematically assess face validity. The staff assessment recommended a more robust logic model substantiated with face validity from the TEP would provide some modest support for a validity claim.

CBE #1425: Measurement of nPCR[3] for Pediatric Hemodialysis Patients [UMICH/CMS]

Specifications

Measure Description: *Percentage of patient months of pediatric (< 18 years old) in-center hemodialysis patients (irrespective of frequency of dialysis) with documented monthly nPCR measurements.*

Staff Preliminary Assessment Rating

Importance: Met

Rationale: This measure meets all criteria for "Met" due to its robust, well-graded evidence base, clear business case, documented performance gap, significant anticipated impact, well-articulated logic model, and is superior to existing measures, making it essential for addressing pediatric dietary nutrition in in-center hemodialysis patients.

Feasibility: Met

Rationale: This maintenance 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 health care system.

Reliability: Met

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

Validity: Not Met

Rationale: The developer conducted person- or episode-level validity testing on key data elements from 300 facilities, showing high agreement and low missing values, but did not perform accountable entity-level validity testing due to small pediatric sample sizes. The developer also did not systematically assess face validity. The absence of material variation of an intermediate outcome measure among entities in the Importance Table (Table 1) also does not support a validity claim. Going forward, a more robust logic model substantiated with face validity from the TEP would provide some modest support for a validity claim.

Equity: Not Met

Rationale: The developer did not address this optional domain.

Use & Usability: Met

Rationale: For maintenance, the measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. The measure also demonstrates a positive trend in performance results, affirming its ongoing usability. The developer reports no unexpected findings.

[3] Normalized Protein Catabolic Rate (nPCR) is a clinical measurement of nutrition used to assess the protein intake of patients undergoing dialysis. It estimates the rate at which protein is being metabolized in the body and is adjusted for the patient's body size.

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
Moving Beyond a Process Measure: Several Advisory Group members said, particularly as the measure is an older one, they hope the developer is devoting resources and moving toward creating a measure with a recommended nPCR range rather than simply having a process measure. Some Advisory Group members did note this was an important first step. The Advisory Group emphasized that the Recommendation Group should take this into consideration.	When this measure was undergoing development 10 to 15 years ago, the TEP was unwilling to set a specific threshold. They agree with the Advisory Group comments but are limited by the current state of evidence as they are not certain they would be able to move a new measure with a threshold through the endorsement process.
Facility Services: An Advisory Group member asked if the measure accounts for the quality of services or supports provided by the facility.	The measure does not require the facility to take any further steps but measuring; however, they expect facilities to do the right thing and use the information to identify patients who are eligible for various programs. The facilities could provide food counseling, and many facilities have an oral supplemental program.
Measure Adherence: An Advisory Group member asked if the developer is trying to achieve 100% reporting rate.	They currently have high rates of reporting, and as this is a process measure, they would like that rate to be quite high.
Importance: An Advisory Group member echoed that this measure is important for pediatric patients as parents are often making decisions based on adult studies.	Not applicable.
Hospital Setting: An Advisory Group member asked if the measure included hospital-based settings.	The measure includes hospital-based settings.
Protein: An Advisory Group member asked if the measure took high amounts of protein into consideration.	That is a separate discussion of measurement, with the consideration of how much protein the pediatric population should be consuming.
Risk Adjustment: An Advisory Group member commented that they believed all the measures developed by University of Michigan and under discussion at this meeting should be risk adjusted.	Battelle staff clarified that process measures are not typically risk adjusted.
Burden: An Advisory Group member asked if monthly is reasonable for the facilities.	Bloodwork is collected as a routine part of care, so the burden is low.

Feedback/Questions	Summary of Developer Response
Validity at Accountable Entity Level: An Advisory Group member asked if there is evidence to support validity at the accountable entity level.	They are working with a small (but vulnerable) population. There are few dialysis facilities only for pediatric populations, so they are unable to come up with additional analyses at this time.
Health Literacy: A patient participant Advisory Group member commented that this measure may be difficult to understand for individuals who do not have a high degree of knowledge related to dialysis or health literacy.	The developer did not respond to this comment.
Composite Measure: Please see description of topic under #0318.	Please see response under #0318.

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

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance	Public Comment	The American Society of Pediatric Nephrology, American Society of Nephrology, and Kidney Care Partners support the endorsement of CBE #1425 as a reasonable reporting measure for assessing pediatric nutrition, acknowledging its limitations, such as the underlying data being primarily linked to adolescents and the measure not being perfect, but emphasizing its importance as a step towards incorporating pediatric-specific growth or nutrition measures in the ESRD QIP.
Probing	Moving Beyond a Process Measure	Advisory Group	The Advisory Group would like the Recommendation Group to consider whether this is still appropriate as a process measure or whether it would be possible to consider a new measure that recommends a range or threshold for nPCR.
Dissenting	Validity Testing	Staff Assessment	The developer conducted person- or episode-level validity testing on key data elements from 300 facilities, showing high agreement and low missing values, but did not perform accountable-entity level validity testing due to small pediatric sample sizes, which is evidence with narrow variation seen in performance gap Table 1. The developer also did not systematically assess face validity. The staff assessment recommended a more robust logic model substantiated with face validity from the TEP would provide some modest support for a validity claim.

CBE #2706: Pediatric Peritoneal Dialysis Adequacy: Achievement of Target Kt/V[4] [UMICH/CMS]

Specifications

Measure Description: *Percentage of pediatric (< 18 years old) peritoneal dialysis patient-months whose delivered peritoneal dialysis dose was a weekly Kt/Vurea >= 1.8 (dialytic + residual)*

Staff Preliminary Assessment Rating

Importance: Not Met but Addressable

Rationale: This maintenance measure includes a 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 peritoneal dialysis in pediatric populations. However, grading was not provided for the clinical guidelines cited, and the provided evidence is fairly dated (potentially due to a lack of updated clinical guidelines).

Feasibility: Met

Rationale: This maintenance 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 health care system.

Reliability: Met

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

Validity: Not Met

Rationale: The developer conducted person- or episode-level validity testing on key data elements from 300 facilities, showing high agreement and low missing values, but did not perform accountable entity-level validity testing due to small pediatric sample sizes and did not systematically assess face validity. The absence of material variation of an intermediate outcome measure among entities in the Importance Table (Table 1) also does not support a validity claim. Going forward, a more robust logic model substantiated with face validity from the TEP would provide some modest support for a validity claim.

Equity: Not Met

Rationale: The developer did not address this optional domain.

Use & Usability: Met

Rationale: For maintenance, the measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. The measure also demonstrates a positive trend in performance results, affirming its ongoing usability. The developer reports no unexpected findings.

[4] Kt/V measures dialysis adequacy.

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
Importance: Several Advisory Group members highlighted the same importance issues as discussed for CBE #1423, including that the population may be small but is among the most vulnerable. The members said the small sample size should not be a deterrent because pediatric dialysis care is often centralized and because the measure can be powered with appropriate statistical methods.	This measure is important and meaningful from patients. On one of their TEPs, patient members voting overwhelmingly to retain Kt/V and said they felt reassured by seeing their Kt/V numbers.
Shared Decision-Making: An Advisory Group member said they liked this measure because it encouraged shared decision-making and could potentially help pediatric patients become involved in their own care as they become older.	Conducting dialysis at home requires buy-in and shared decision-making.
Health Literacy: A patient participant Advisory Group member commented that this measure may be difficult to understand for individuals who do not have a high degree of knowledge related to dialysis or health literacy.	The developer did not respond to this comment.
Composite Measure: Please see description of topic under #0318.	Please see response under #0318.
Risk Adjustment: An Advisory Group member commented that they believed all the measures developed by University of Michigan and under discussion at this meeting should be risk adjusted.	Please see response under #0318.

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

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Supportive	Importance	Advisory Group; Public Comment	The Advisory Group highlighted the importance of pediatric measures and that small sample sizes should not be a deterrent. The American Society of Pediatric Nephrology, American Society of Nephrology, and Kidney Care Partners support the continued endorsement of CBE #2706 emphasizing its alignment with KDOQI

Discussion Category	Key Themes	Source of Comment	Summary of Comments
			guidelines and its importance for accountability in treating pediatric patients, while recommending maintaining the pediatric-specific Kt/V goal of 1.7 rather than aligning with adult standards, despite the measure potentially being topped out.
Dissenting	Evidence	Staff Assessment	The staff assessment identified that grading was not provided for the clinical guidelines cited, and the provided evidence is fairly dated (potentially due to a lack of updated clinical guidelines).
	Validity Testing and Risk Adjustment	Advisory Group; Staff Assessment	The Advisory Group questioned whether the measure should be risk adjusted. While a risk adjustment evaluation is not required for intermediate outcome measures, a discussion of risk adjustment would strengthen the submission. The developer conducted person- or episode-level validity testing on key data elements from 300 facilities, showing high agreement and low missing values, but did not perform accountable-entity level validity testing due to small pediatric sample sizes, which is evidence with narrow variation seen in performance gap Table 1. The developer also did not systematically assess face validity. The staff assessment recommended a more robust logic model substantiated with face validity from the TEP would provide some modest support for a validity claim.

CBE #3502e: Hybrid Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure with Claims and Electronic Health Record Data [Yale CORE/CMS]

Specifications

Measure Description: *Hybrid Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure with Claims and Electronic Health Record Data measure estimates a hospital-level 30-day risk-standardized mortality rate (RSMR), defined as death from any cause within 30 days after the index admission date for Medicare fee-for-service and Medicare Advantage patients who are between the ages of 65 and 94.*

Index admissions are assigned to one of 15 clinically cohesive and mutually exclusive divisions: six surgical divisions and nine non-surgical divisions, based on the reason for hospitalization. The surgical divisions are: Surgical Cancer (includes a surgical procedure and a principal discharge diagnosis code of cancer), Cardiothoracic Surgery, General Surgery, Neurosurgery, Orthopedic Surgery, and Other Surgical Procedures. The non-surgical divisions are: Cancer, Cardiac, Gastrointestinal, Infectious Disease, Neurology, Orthopedic, Pulmonary, Renal, Other Conditions. The final measure score (a single risk-standardized mortality rate) is calculated from the results of these 15 different divisions, modeled separately. Variables from administrative claims and electronic health records are used for risk adjustment.

Staff Preliminary Assessment Rating

Importance: Not Met but Addressable

Rationale: This maintenance measure addresses a wider patient group not covered by existing mortality metrics, linking various hospital processes to better health outcomes and lower mortality rates. Analysis from Medicare FFS and MA data reveals less variation observed in hybrid datasets, possibly reflecting higher performance reporting. The measure was developed with patient input, enhancing its relevance and importance to the patient community.

Feasibility: Met

Rationale: The measure utilizes both claims and EHR data. Feedback from hospitals has led to several adaptations, such as revising data standards to address issues like unusable data units and low capture rates. Continuous updates, influenced by hospital feedback, include expanded data-collection windows and optional reporting of certain lab values to improve data accuracy and completeness. The measure utilizes existing EHR systems with minimal estimated costs and disruption to clinical workflows.

Reliability: Met

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

Validity: Met

Rationale: The developer conducted face validity and person- or episode-level validity testing, showing agreement on the measure's ability to distinguish facility quality but with insufficient consensus data, and found varying agreement between EHR-based and chart-abstracted data, while accountable-entity validity testing showed expected correlations; statistical risk adjustment was performed without including social risk factors due to minimal impact, achieving c-statistics

of 0.75-0.90, but causal claims remain prone to bias without additional mechanism studies. Going forward, additional studies that either rule out potential confounding (in addition to risk adjustment) or describe features of potential mechanisms will strengthen causal claims. The inclusion of a study on nurse-staffing ratios and face validity strengthens causal claims.

Equity: Not Met but Addressable

Rationale: The developer evaluated two social risk factors, DE and ADI. Despite higher unadjusted rates of adverse outcomes among patients with high social risk, the empirical results showed minimal impact on adjusted measure scores, with correlation coefficients nearly identical (0.999) and minimal median changes in risk-standardized mortality rates. Consequently, the developer decided not to adjust the measure for these factors, supporting robust calibration and the measure's ability to evaluate hospital performance fairly across varying social risk profiles. However, the limitation noted is the lack of stratification by social risk factors, which could provide deeper insights into disparities in outcomes.

Use & Usability: Not Met but Addressable

Rationale: The updated Hybrid HWM measure is set to be included in the HIQR program, replacing the previous claims-only measure, with significant updates such as the inclusion of MA. Hospitals are encouraged to adopt evidence-based actions to reduce 30-day mortality rates. Feedback mechanisms through the Q&A tool on QualityNet and the Annual Updates Process have led to active stakeholder engagement, resulting in updates to accommodate more data units and extend the CCDE lookback period, among other modifications. Despite these advancements, challenges with EHR data submission and meeting CMS programmatic data reporting thresholds were noted. The measure's initial voluntary reporting in 2024 saw limited participation, which poses challenges in generalizing improvements, but no unintended impacts on patient care were reported.

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
Equity: An Advisory Group member stated that they appreciated that the developer addressed equity even though that domain is still optional.	Not applicable.
Overlap in Divisions: An Advisory Group member asked what happens if an admission can be applied to multiple divisions.	If there are multiple conditions, they use the condition associated with the most billing codes from the claims data.
Claims Data: An Advisory Group member asked how the measure takes into consideration how claims data are geared to provide hospitals with the best reimbursement.	They have a hierarchy that was created by input from clinicians and consider the clinical status when the patient arrives, and risk adjust for case mix.
Feasibility: An Advisory Group member expressed concern that this measure would be burdensome for hospitals to track.	The measure uses the Medicare beneficiary file, so there is no burden for hospitals.
Importance: A patient member of the Advisory Group found this to be an important measure.	Not applicable.
Sepsis Discharge Planning: An Advisory Group member asked if discharge planning was included for sepsis.	Discharge planning is outside the scope of the measure.
Usefulness of Composite Measures: The Advisory Group discussed whether a composite measure such as this is helpful for clinicians and facilities. One Advisory Group member voiced that a composite measure is preferable to having a multitude of measures created by individual societies and that they keep clinicians focused and motivated. A few other Advisory Group members disagreed, saying that it can be difficult to create an action or improvement plan based on the information provided by the composite measure.	Not applicable.
Improvements: An Advisory Group member asked if the developer has seen improvement since implementation.	The measure has not been implemented yet. It was in voluntary reporting. They have seen some improvement in readmission rates.
Exclusions: An Advisory Group member asked if palliative care and hospice care are excluded from the measure.	Hospice care, palliative care, and diagnoses where the hospital has limited ability to impact a patient's ability to survive are excluded. In addition, the patient would be excluded if they are in hospice 12

Feedback/Questions	Summary of Developer Response
	months prior to admission or admitted to hospice 2 days after admission.
Scores: An Advisory Group member asked if facilities are given a single, final score or if they are given scores that allow to see where gap areas are.	Facilities received condition-specific scores by division.
Trauma Surgery versus General Surgery: An Advisory Group member asked if trauma surgery patients are separated from general surgery.	Patients who are primarily treated for injury, burns, intracranial injury, spinal cord injury, skull and face fracture, or open wounds of the head, neck, spine, and trunk are excluded from the measure.
Parameter Estimates: An Advisory Group member requested the developer's parameter estimates.	Battelle staff referred the Advisory Group to the Supplemental Attachment section 7.1, where additional measure information can be found for the measure.
COVID-19 Exclusions: An Advisory Group member asked why COVID-19 was excluded rather than being risk-adjusted.	The developer did not respond to this comment.
Feature-Selection Methods: An Advisory Group member expressed that stepwise regression as the feature-selection methodology is outdated and urged the developer to consider machine learning methods (e.g., least absolute shrinkage and selection operator [LASSO]) in future updates.	The developer did not respond to this comment.
Health Literacy: A patient Advisory Group member said they believed this measure would be relatively comprehensible for a broad audience with varied levels of health literacy.	Not applicable.

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

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Mixed	Usability	Advisory Group	The Advisory Group disagreed with how useful the information coming out of this measure would be for facilities and clinicians to make improvements. One Advisory Group member voiced support, but others disagreed.
Dissenting	Feasibility	Advisory Group; Staff Assessment; Public Comment	The staff assessment identified that challenges with EHR data submission and meeting CMS programmatic data reporting thresholds were noted. The measure's initial voluntary reporting in 2024 saw limited participation, which poses challenges in generalizing improvements. A public comment from the American Medical Association (AMA) noted the challenges with data collection and submission of measures that

Discussion Category	Key Themes	Source of Comment	Summary of Comments
			leveraged data from EHR systems and that the current measure specifications do not align with current workflows.
	Measure Impact and Gap	Staff Assessment	Analysis from Medicare FFS and MA data reveals less variation observed in hybrid datasets, possibly reflecting higher performance reporting.
Probing	Equity	Staff Assessment	For this optional domain, the developer evaluated social risk factors DE and ADI, finding minimal impact on adjusted measure scores and nearly identical correlation coefficients, leading to the decision not to adjust for these factors, though the lack of stratification by social risk factors limits insights into outcome disparities.

CBE #4595: Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Acute Ischemic Stroke Hospitalization with Claims-Based Risk Adjustment for Stroke Severity [Yale CORE/CMS]

Specifications

Measure Description: *The measure estimates the hospital-level, risk-standardized mortality rate (RSMR) for Medicare patients (Fee-for-Service [FFS] and Medicare Advantage [MA]) discharged from the hospital with a principal discharge diagnosis of acute ischemic stroke. The outcome is all-cause 30-day mortality, defined as death from any cause within 30 days of the index admission date, including in-hospital death, for stroke patients. The measure includes the National Institutes of Health (NIH) Stroke Scale as an assessment of stroke severity upon admission in the risk-adjustment model.*

Staff Preliminary Assessment Rating

Importance: Met

Rationale: This new measure, utilizing Medicare FFS and MA data, aims to replace the current claims-only measure in the Hospital Inpatient Quality Reporting program. It highlights that hospitals employing evidence-based actions and participating in programs like Get With The Guidelines often see improved outcomes through integrated care models and multidisciplinary teams. However, the measure's logic model diagram includes generalized hospital inputs that may not be fully supported by the evidence provided and lacks some details such as high service volume and staff training. The measure is intended to enhance transparency and accountability in stroke care, informing both patients about hospital performance.

Feasibility: Met

Rationale: The measure leverages routinely generated claims data, requiring no additional data collection by hospitals, thus minimizing reporting burdens and implementation challenges. Managed and processed securely by CMS, this measure ensures patient confidentiality and data accuracy, with health care facilities reporting no significant concerns regarding its implementation.

Reliability: Not Met but Addressable

Rationale: There are potential issues with the accuracy of the results using the current reliability metrics. However, the identified limitations are deemed addressable, as the developer may consider performing additional reliability testing such as split-half reliability. By addressing this issue, there is potential to enhance the reliability.

Recommended Action to Address Reliability Concerns:

- Refine Reliability Testing Methods: Adjust the reliability testing methods or analytic approaches to improve the assessment the reliability of the final measure.

Validity: Met

Rationale: Because this is a new measure, person- or episode-level validity is sufficient, although this measure is new only with respect to the risk adjustment model.

The risk adjustment methods used are appropriate and demonstrate variation in the prevalence of risk factors across measured entities, contribute to unique variation in the outcome, and show the impact of risk adjustment for providers at high or low extremes of risk. The model performance is acceptable.

Equity: Not Met but Addressable

Rationale: The developer evaluated two social risk factors, DE and ADI. Despite higher unadjusted rates of adverse outcomes among patients with high social risk, the empirical results showed minimal impact on adjusted measure scores. Consequently, CMS decided not to adjust the measure for these factors. However, the limitation noted is the lack of stratification by social risk factors, which could provide deeper insights into disparities in outcomes.

Use & Usability: Met

Rationale: This new measure, planned for use in the HIQR program, incorporates significant improvements based on extensive stakeholder feedback and evidence-based practices. Hospitals are encouraged to engage in telestroke networks, maintain specialized stroke care teams, and enhance staff training to reduce mortality rates, with additional support from detailed CMS reports for ongoing quality improvement. The measure, continuously refined through expert input and public comments, now includes risk adjustments for stroke severity and an expanded cohort to include MA beneficiaries.

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
Unintended Consequences: An Advisory Group member requested more information on potential unintended consequences and mitigation strategies.	They did not identify any unintended consequences during measure development or model testing. They are committed to monitoring the measure's use and assessing potential unintended consequences over time, such as inappropriate shifting of care, and increased patient morbidity and mortality. This measure complements other existing quality measures used within quality improvement programs, noting that these measures are limited to process measures and measures of functional outcomes that are not publicly reported.±
Unintended Consequences on Rural/Under-Resourced Communities: An Advisory Group member asked if this measure negatively impacts facilities that are not stroke centers and that serve rural or under-resourced communities.	The measure helps hospitals improve, whether or not they are stroke centers, by allowing them to track their mortality rates. Hospitals that admit fewer than 25 stroke patients during the performance period do not have their performance scores publicly reported. These hospitals do, however, receive, confidentially, their measure score, in addition to patient-level information about the stroke patients that were admitted to their hospital, to support quality improvement efforts.±
Unintended Consequences Related to Transfer of Patients: An Advisory Group member asked if a facility may be reluctant to accept a patient if they believe they would be penalized by the measure.	Once a patient is admitted, the mortality is attributed to the original facility. If a patient is seen in the emergency department (ED) but not admitted, that is not counted as an admission.
Separate from CBE #3502e: An Advisory Group member asked why this measure is separate from measure CBE #3502e.	This measure was developed prior to the composite and helps with processes directly tied with outcomes for ischemic stroke, meaning that it serves a different purpose.
Usefulness: An Advisory Group member said that because the data are significantly delayed, facilities struggle to identify opportunities for improvement.	The developer did not respond to this comment.
Imputing: An Advisory Group member remarked on how the developer's scale goes from 1-42 but they used a zero for missing data. They asked what the clinical relevance was of not saying the scale went from 0-42.	They tested multiple imputation methods. They used zero to encourage hospitals to report the National Institutes of Health (NIH) Stroke Scale, which was not as widely reported previously. The zero is not intended to represent an actual value of zero.

Feedback/Questions	Summary of Developer Response
New Measure: An Advisory Group member asked why this was considered a new measure rather than an update to an existing measure used in a program.	Battelle staff clarified that if something is considered a "new" measure, that means it is new to the endorsement process. In response to stakeholder feedback, the developer added the NIH Stroke Scale and incorporated Medicare Advantage (MA) patients into the target population.
Timeframe: An Advisory Group member commented that 30 days might not be long enough.	It is difficult to attribute mortality to the hospital when going longer than 30 days. They said 30 days is reasonable and aligns with other quality measures.
Feasibility: An Advisory Group member asked how mortality data is collected for this measure, as it can be difficult to collect from private insurers.	This measure applies to MA or fee-for-service (FFS). The Social Security Administration gives that information to CMS, and then the developer has access to a file with the information in it.
Burden: An Advisory Group member commented that some facilities may decrease mortality but increase burden on the patient.	Measuring mortality separately is important; without mortality, all that remains is process measures and functional outcomes. A hospital may do well on functional outcomes but still have a high mortality rate.
Feature-Selection Methods: An Advisory Group member expressed that stepwise regression as the feature selection methodology is outdated and urged the developer to consider machine learning methods (e.g., least absolute shrinkage and selection operator [LASSO]) in future updates.	The developer did not respond to this comment.
Health Literacy: A patient participant Advisory Group member said they believed this measure would be relatively comprehensible for a broad audience with varied levels of health literacy.	Not applicable.
Underlying Risk: An Advisory Group member asked what is responsible for the underlying risk of stroke for patients in the cohort.	They empirically selected risk variables in claims data and, following that, added additional variables, including the NIH Stroke Scale (NIHSS) to the list of risk variables. The NIHSS is one of more than 95 variables in the risk model. The NIHSS ranges from 0-42; the developer uses the numerical NIHSS score in the regression model. For every unit increase in NIHSS score the odds of death increase by 4%. Risk variables (ICD-10 codes) with the highest odds ratios during the index admission include: • Secondary malignant neoplasm of liver and intrahepatic bile duct (OR of 6.51); • Compression of brain (OR of 2.51); • NonST elevation (NSTEMI) myocardial infarction (OR of 2.03); • Acute respiratory failure, unspecified whether with hypoxia or hypercapnia (OR of 3.14); • Acute respiratory failure with hypoxia (OR of 2.26);

Feedback/Questions	Summary of Developer Response
	• Encounter for palliative care (OR of 22.81). See Table 9 in the "All Figures and Tables Stroke Mortality" attachment for more details.±

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

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

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Dissenting	Reliability Testing	Staff Assessment; Public Comment	There are potential issues with the accuracy of the results using the current reliability metrics. However, the identified limitations are deemed addressable, as the developer may consider performing additional reliability testing such as split-half reliability. In addition, the American Medical Association request that a case minimum of 25 individuals be established for this measure to ensure a minimum reliability close to 0.7.
	Usability	Advisory Group	The Advisory Group wondered whether facilities or clinicians could use the information from this measure to drive improvements.
	Unintended Consequences	Advisory Group	The Advisory Group discussed potential unintended consequences that could stem from this measure, including its impact on under-resourced and rural communities.
Probing	Equity	Staff Assessment	For this optional domain, the developer evaluated social risk factors DE and ADI, finding minimal impact on adjusted measure scores, leading to the decision not to adjust for these factors, though the lack of stratification by social risk factors limits insights into outcome disparities.

CBE #0531: Patient Safety Indicator (PSI) 90: Patient Safety and Adverse Events Composite [Mathematica/CMS]

Specifications

Measure Description: *PSI 90 is a composite of ten adverse event indicators that summarizes hospitals' performance on patient safety for the CMS Medicare fee-for-service population. The timeframe used in the CMS Hospital Acquired Conditions Reduction Program (HACRP) and CareCompare public reporting are set within the Inpatient Prospective Payment Systems (IPPS) Final Rule annually. Typically, the performance periods use multiple months of claims data.*

Staff Preliminary Assessment Rating

Importance: Not Met but Addressable

Rationale: The measure is rated as "Not Met, but Addressable" due to incomplete information on grading of systematic review evidence and guidelines for most component measures. It is also not clear from the submission which component measures are currently endorsed. A more thorough presentation of grading of evidence, and potentially more robust evidence for a few components, could elevate its importance.

Feasibility: Met

Rationale: This maintenance measure meets all criteria for "Met" due to its clear and implementable data collection strategy, and transparent handling of licensing and fees, ensuring practical implementation within the health care system.

Reliability: Met

Rationale: The developer conducted split-half reliability testing at the accountable-entity level. A PSI composite score was calculated for each split within each hospital and a random effects model was fit to calculate the intra-class correlation coefficient (ICC). The Spearman-Brown formula applied. More than 70% of accountable entities meet the expected threshold of 0.6.

Validity: Met

Rationale: The validity testing results support a moderate inference of validity for the measure, confirming that the measure accurately reflects performance on patient safety and can distinguish good from poor performance. The risk adjustment methods used by the component measures are appropriate and demonstrate variation in the prevalence of risk factors across measured entities, contribute to unique variation in the outcome, and show the impact of risk adjustment for providers at high or low extremes of risk. The model performance is acceptable.

Equity: Not Met but Addressable

Rationale: While the measure partially addresses equity in health care outcomes for each component measure, they could 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: For maintenance, the measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. The developer provides evidence of opportunities for entities to improve performance, through their evidence review, through annual reports to entities, and through the availability of an AHRQ toolkit for performance improvement. Despite no large improvements in performance in the measure performance over time, the developer notes that large changes in performance are not expected in composite measures, and also notes that three component measures did show improvement in recent years. The developer reports no unexpected findings.

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
Importance: Several Advisory Group members, including patient participants, expressed how important this measure is, with some calling it the most important measure of the day. They commented that this measure creates better outcomes for patients and holds entities accountable. They noted that many of the complications are preventable, and that the measure brings transparency and focus to patient safety and adverse events.	Not applicable.
Time Lag: Advisory Group members asked, from the time when it is reported, how long it takes for the information to appear on Care Compare. They also asked if that time frame could be shortened	It takes 2 years due to how claims are submitted.
Population Expansions: An Advisory Group member asked if this measure could be expanded to include pediatrics and to include medication harms.	They, and CMS, agree that these issues should be looked at in pediatrics. They are currently making sure the measure components are appropriate for a pediatric population and that there would be enough cases to include. Medication harms are handled separately from this measure.
Health Literacy: A patient participant Advisory Group member said they believed this measure would be relatively comprehensible for a broad audience with varied levels of health literacy.	Not applicable.

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

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Mixed	Importance	Advisory Group; Staff Assessment	The Advisory Group, and particularly the patient members, highlighted the importance of this measure because it creates accountability and transparency and is meaningful to patients. However, the staff assessment rates Importance as "Not Met, but Addressable" due to incomplete evidence grading and unclear

Discussion Category	Key Themes	Source of Comment	Summary of Comments
			endorsement of component measures, with a need for more thorough evidence presentation to enhance its importance.
Dissenting	Reliability	Public Comment	The AMA questions the usefulness of the composite measure for accountability and improvement due to poor reliability, low ICC, and reliance on delayed administrative claims data, suggesting the committee discuss requiring a case minimum to potentially improve reliability.
	Equity	Staff Assessment	For this optional domain, the measure partially addresses equity in health care outcomes, but more information on methods and interpretation of disparities could enhance understanding and improvement efforts.

CBE #0753: 30-Day Post-Operative Colon Surgery (COLO) and Abdominal Hysterectomy (HYST) Surgical Site Infection (SSI) Standardized Infection Ratio (SIR) [Centers for Disease Control and Prevention (CDC)]

Specifications

Measure Description: *Annual risk-adjusted standardized infection ratio (SIR) of observed over predicted deep incisional primary and organ/space surgical site infections (SSIs), over a 30-day post-operative surveillance period, among hospitalized adults who are >=18 year of age with a date of admission and date of discharge that are different calendar days, and the patient underwent a colon surgery (COLO) or abdominal hysterectomy (HYST) at an acute care hospital or oncology hospital. The 30-day postoperative surveillance period includes SSIs detected upon admission to the facility or a readmission to the same facility or a different facility (other than where the procedure was performed) and via post-discharge surveillance.*

Staff Preliminary Assessment Rating

Importance: Met

Rationale: This new measure meets all criteria for “Met” due to its robust, well-graded evidence base, clear business case, documented performance gap, and significant anticipated impact, making it essential for addressing SSIs for hysterectomies and colon surgeries.

Feasibility: Met

Rationale: This maintenance measure meets the criteria for “Met” due to the collection of all data elements during usual patient care, and no implementation barriers related to licensing or fees.

Reliability: Met

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

Validity: Met

Rationale: The validity testing results support a 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. The risk adjustment methods used are appropriate and demonstrate variation in the prevalence of risk factors across measured entities, contribute to unique variation in the outcome, and show the impact of risk adjustment for providers at high or low extremes of risk. The model performance is acceptable.

Equity: Not Met

Rationale: The developer did not address this optional domain.

Use & Usability: Not Met but Addressable

Rationale: For maintenance, the measure is actively used in multiple accountability applications, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. The measure also demonstrates a positive trend in performance results for colon surgeries, affirming its ongoing usability for that purpose. Improvement in hysterectomy SSIs

was not clear, and this submission would be strengthened by providing a rationale for the lack of clear improvement. The developer reports no unexpected findings.

Public Comment

Number of Comments Received During the Public Comment Period: 2

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
Accountable Entity Actions: An Advisory Group member asked what actions the accountable entities can take to improve their scores.	Facilities can take many actions to improve their scores, depending on what systems they already have in place. The Infectious Diseases Society of America (IDSA), Society for Healthcare Epidemiology of America (SHEA), and the Association for Professionals in Infection Control and Epidemiology (APIC) have practice recommendations and strategies to prevent SSIs. Some of these practices include but are not limited to: • Antimicrobial prophylaxis • Preoperative glycemic control • Normothermia maintenance • Sterile technique • Avoidance of preoperative shaving. See questions 2.1 and 2.2 in their measure submission for more information.
Measure Importance: A patient member of the Advisory Group noted that this measure is important, highlighting that one study found 28% of these surgeries resulted in readmission because of infection. Another patient agreed that the measure is important, sharing that a few friends had hysterectomies, and one had a hard time recuperating.	Not applicable.
Unresolved Issues: An Advisory Group member asked why surgical attire is considered an unresolved issue. A patient participant Advisory Group member shared a personal experience where a new protocol was being implemented during surgery. She noted that unresolved issues could undermine trust and safety in the health care setting.	They were citing 2022 updated practice recommendations from IDSA, SHEA, and APIC, which states, "Although there are longstanding traditions and opinions regarding surgical attire in the operating room, no strong evidence exists for many of them" (p. 29). That report was noting unresolved issues related to the grading of clinical practice guidelines and was not specific to the measure. They clarified that

Feedback/Questions	Summary of Developer Response
	they were not involved the drafting or maintaining of these recommendations.±
Risk-Adjustment Model: The Advisory Group talked about several facets of risk adjustment, including: • **Facility-Level Unintended Consequences:** A few Advisory Group members were concerned that the measure would unfairly penalize certain types of facilities. One Advisory Group member said they believed the measure would reward large non-trauma hospitals and penalize smaller hospitals that are in rural or under-resourced areas because the measure uses hospital/facility-level characteristics in the risk model. They noted that it was not appropriate to use hospital as a fixed effect in the risk-adjustment model. • **Patient-Level Risk Adjustment:** Several Advisory Group members said they believed that the measure should be risk adjusted by patient-level characteristics. An Advisory Group member pointed out that if the measure was risk adjusted by patient-level characteristics and did not use hospital as a fixed effect, the measure would not reward large hospitals. • **Hierarchical Clustering:** An Advisory Group member stated they believed the model should have hierarchical clustering. • **Feature-Selection Methods:** An Advisory Group member stated that the feature selection methodology used in risk adjustment of this measure is outdated and biased. They suggested manuscripts (2019 in The American Statistician and 2020 in the International Journal of Epidemiology) on stepwise regression creating unfair risk-adjustment models for the Recommendation Group. • **Moderate C Statistic:** While a few Advisory Group members criticized the model for having a moderate C statistic of 0.6, a patient participant pointed out that while the number is smaller, the committee at least knows they are dealing with truthful numbers and it something that can be improved upon. Overall, a few Advisory Group members wished they had been given more methodology information by the developer in the submission materials.	The measure is risk adjusted at the procedure level, which takes into account whether or not a case is a trauma case. They use a data-driven approach, and their evidence shows that even though they are already incorporating patient-level data at the procedure level, there is still value in including facility-level risk factors. Among those, they had found evidence of increased SSI in facilities with a larger bed count. This is a fixed-effect model, not a mixed model, which enables them to observe rates for the baseline year and then evaluate the facilities in the following years. Further written information as to why they used a fixed-effect model, which can be found in Appendix A. The risk-adjustment model has been updated since initial endorsement. The new model includes sex at birth, BMI, and age; it does not currently include race, ethnicity, and insurance status because they are cognizant of the burden of reporting on facilities. However, as the National Healthcare Safety Network (NHSN) may soon require that information to be reported, those factors may be incorporated into the measure by the next time it returns for maintenance. Their C statistics are 0.635 (colon) and 0.623 (hysterectomy). Improving the C statistic would require heavier data collection and potentially yield marginal increases. They outlined that several other performance measures have C statistics within this range.±
Definition of a "Large" Facility: An Advisory Group member expressed concern with how the developer had defined a "large"	Their process for determining in cutoff points is data driven and found evidence of increased SSI in facilities with a larger bed count. In this

Feedback/Questions	Summary of Developer Response
facility based on the number of beds in the facility and that number was not chosen through analysis.	specific case, they observed significant differences in SSI risk at facilities that had 319 beds or more.±
Academic versus in Practice: An Advisory Group member noted that in academic studies, there seem to be many ways to control for infection that seem to be less effective in the real world. They wondered if that was a nature of complexity (e.g., of infection prevention bundles), lack of resources, or if focus would help reduce the number. They suggested the Recommendation Group consider why there hasn't been more improvement.	Not applicable.
Health Literacy: A patient participant Advisory Group member said they believed this measure would be relatively comprehensible for a broad audience with varied levels of health literacy.	The developer did not respond to this comment.
Trauma Cases: An Advisory Group member asked why the measure captures trauma cases and whether trauma is considered in risk adjustment.	The NHSN Patient Safety Component SSI Protocol is available to all acute-care/critical-access hospitals including those performing surgeries related to trauma, which are relatively low volume. The 30-day model uses trauma, among other factors, as a predictor of infection following COLO. It may not be possible to control for trauma itself, but surveillance of these events is a critical part of developing effective strategies for prevention of SSIs.±

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

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

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Mixed	Risk Adjustment	Advisory Group; Staff Assessment	The Advisory Group talked at length about the risk adjustment for this measure, including whether facility-level and patient-level characteristics were appropriately included, whether hospital should be a fixed effect, whether hierarchical clustering should be done, and whether the feature selection method used is outdated and biased. The Advisory Group also felt the submission materials could have included more information and noted the C statistic was moderate. The staff assessment noted that the risk adjustment methods used are appropriate and demonstrate variation in the prevalence of risk factors across measured entities, contribute to unique variation in the

Discussion Category	Key Themes	Source of Comment	Summary of Comments
			outcome, and show the impact of risk adjustment for providers at high or low extremes of risk. The model performance is acceptable.
Dissenting	Exclusions	Public Comment	The Memorial Hermann Texas Medical Center and McGovern Medical School UTHealth Houston urge the exclusion of trauma-related colon surgeries from SSI surveillance metrics due to their inherent high risk and impact on standardized infection ratios, advocating for separate categorization to ensure fair performance evaluation and targeted improvement. The American Medical Association supports this exclusion due to differing factors from elective cases.
	Usability	Staff Assessment	Improvement in hysterectomy SSIs was not clear, and this submission would be strengthened by providing a rationale for the lack of clear improvement.
	Minimum Case Volume	Public Comment	The American Medical Association calls for a case minimum to improve measure reliability.
Supportive	Importance	Advisory Group	A few patient participants noted the importance of this measure, highlighting that evidence showed a 28% readmission rate due to infection following these surgeries, that there is opportunity to improve, and they believed this measure would be relatively comprehensible for a broad audience with various levels of health literacy.
Probing	Practical Improvement	Advisory Group	The Advisory Group asked the Recommendation Group to consider the practicalities that go into reducing infection rates in the real world and the complexities, resources, and dedication to time and attention that may go into that process.

CBE #3309 Risk-Standardized Survival Rate (RSSR) for In-Hospital Cardiac Arrest [AHA]

Specifications

Measure Description: *This measure estimates a hospital-level risk standardized survival rate (RSSR) for adult patients aged 18 years and older who experience an in-hospital cardiac arrest.*

Staff Preliminary Assessment Rating

Importance: Not Met but Addressable

Rationale: The measure is rated as “Not Met but Addressable” due to a potentially incomplete evidence review. Enhancements in the evidence, including a discussion of guidelines (if applicable) could elevate its importance.

Feasibility: Met

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

Reliability: Met

Rationale: The results demonstrate sufficient reliability at the patient or encounter and accountable entity levels.

Validity: Met

Rationale: The validity testing results support a 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.

The risk adjustment methods used are appropriate and demonstrate variation in the prevalence of risk factors across measured entities, contribute to unique variation in the outcome, and show the impact of risk adjustment for providers at high or low extremes of risk. The model performance is acceptable.

Equity: Not Met

Rationale: The developer did not address this optional domain.

Use & Usability: Not Met but Addressable

Rationale: The measure is actively used in at least one accountability application and the developer identifies several ways entities can improve performance. There has also been improvement over the brief period since the measure was re-derived to address the effects of the COVID pandemic on the measure. However, the submission could be improved with additional information regarding the approach to collecting and responding to feedback.

Public Comment

This measure did not receive any comments during the public comment period.

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Risk-Adjustment and Stratification: A few Advisory Group members asked for clarification on why the developer chose not to include race/ethnicity, sex, or age in the risk-adjustment model. One Advisory Group member commented making decisions about the risk modeling always includes trade-offs. Another Advisory Group member noted that the risk-adjustment model rates were good but asked for the parameter estimates and noted that the stepwise regression as the feature selection methodology is outdated and urged the developer to consider machine learning methods (e.g., least absolute shrinkage and selection operator [LASSO]) in future updates." One Advisory Group member noted there may be a different set of underlying conditions or survival rates based on patient age and asked if the developer has seen any performance differences related to age during the last 5 years of measure implementation. A patient participant expressed the importance of considering a wider age range in the measure, as both younger individuals (e.g., those with birth defects or rare diseases) and older individuals with chronic conditions are at risk. Advisory Group members expressed interest in seeing measure results stratified by age, race, ethnicity, and sex. The Advisory Group suggested the Recommendation Group consider whether differences in these factors mean differences in ways that patients are treated and monitored (inequities) and how that information would be important to uncover.	They are aware of disparities in care. In consultation with multiple experts, they intentionally did not include these variable as they did not wish to mask any disparities in care with risk adjustment. For example, they analyzed registry data and found that hospitals with a higher proportion of Black patients had lower survival rates compared to those with fewer Black patients. Social risk factors are more relevant to long-term survival and recovery (e.g., access to care and follow-up treatment) as opposed to short-term inpatient survival outcomes (the measure focus). **Staff Note:** In the developer's submission, per Table 5 in the 7.1 Supplemental Attachment, the developer did include age in the final risk-adjustment model.
Missing Data: One Advisory Group member asked about how much missingness there is among these rates; another asked what the developer did with missing race data.	Missingness is fairly minimal because the measure is voluntary to report. They did not include any missing race data to ensure they did not affect the model.
Improvements: An Advisory Group member asked if the developer had information on improvements over time.	They recently rederived the model using 2022 and 2023 data and have been able to demonstrate some improvement with the new model.

Feedback/Questions	Summary of Developer Response
Health Literacy: A patient participant said they believed this measure would be relatively comprehensible for a broad audience with varied levels of health literacy.	The developer did not respond to this comment.

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

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Dissenting	Importance	Staff Assessment	The Importance domain is rated as "Not Met but Addressable" due to potentially incomplete evidence review and insufficient patient input.
	Usability	Staff Assessment	The approach to collecting and responding to feedback is not clear. In addition, there may not be a clear positive trend in performance scores.
Probing	Risk Adjustment and Uncovering Inequity	Advisory Group	The Advisory Group discussed whether the measure should include patient-level risk factors. They highlighted that the Recommendation Group should consider the importance of whether the measure might incidentally uncover inequities in the way that certain groups of patients are treated.

CBE #4580: Composite measure for the quality of care provided to patients undergoing percutaneous coronary interventions (PCI) [ACC]

Specifications

Measure Description: *This is a weighted composite measure comprised of six component measures: three all-cause risk standardized outcome measures on all-cause mortality, bleeding, acute kidney injury and three process measures focused on discharge on guideline directed medical therapy, referral to a cardiac rehabilitation program and PCI performed within ninety minutes of symptoms for patients with acute myocardial infarctions. The target population includes adults (age 18 and greater) undergoing percutaneous coronary interventions. The timeframe for reporting will be a rolling four quarters.*

Staff Preliminary Assessment Rating

Importance: Not Met but Addressable

Rationale: The measure is rated as “Not Met but Addressable” due to a sparse logic model and potentially insufficient patient input. Enhancements in the logic model and more robust patient input could elevate its importance.

Feasibility: Not Met but Addressable

Rationale: The measure is rated as “Not Met but Addressable” due to incomplete discussion of the feasibility assessment performed, and lack of clarity regarding the requirements for participating in the registry to report the measure.

Reliability: Not Met but Addressable

Rationale: The current reliability metrics do not meet the established thresholds, indicating potential issues with the consistency and accuracy of the results across different settings and populations. However, the identified limitation is deemed addressable, as the developer may consider identifying sources of prior evidence for the additional 16 data elements. By addressing this issue, there is potential to enhance the reliability.

Validity: Met

Rationale: As a new measure the person- or encounter-level validity evidence is sufficient, and the weighting scheme sufficiently justified. Going forward the accountable-entity validity will need to be addressed.

The risk adjustment methods used are appropriate and demonstrate variation in the prevalence of risk factors across measured entities, contribute to unique variation in the outcome, and show the impact of risk adjustment for providers at high or low extremes of risk. The model performance is acceptable.

Equity: Not Met but Addressable

Rationale: The rating for Equity is “Not Met but Addressable.” While the submission partially addresses equity, we recommend the developer perform significance testing and provide interpretation of the results, which would include how the results relate to the evidence prior, any limitations of the results, and the potential impact of these differences on the identified

subpopulations. Equity content can be found in Tables 22-25 and Figures 11-14 on pp. 34-40 of their supplemental attachment.

Use & Usability: Met

Rationale: The measure is actively used in at least one accountability application, with a clear feedback approach that allows for continuous updates based on stakeholder feedback. Trend in measure performance results was not reported due to lack of data from the period the measure has been in use. The developer reports no unexpected findings or unexpected adverse findings.

Public Comment

This measure did not receive any comments during the public comment period.

Advisory Group Feedback

Feedback/Questions	Summary of Developer Response
Weighting: A few Advisory Group members expressed concern or asked for more information about how the components are weighted within the composite measure. One Advisory Group member felt that the components should be weighted more equitably, given that they all reflect processes that have been shown to improve patient outcomes.	Five different weighting scenarios were considered. In the final weighting decision, outcome measures combined account for 75% of the composite weight with process measures accounting for 25%. The highest weight was placed on risk of in-hospital mortality (35%). The weighting of the measure was reviewed by the ACC's Metrics and Reporting Methodology (MRM) sub-committee and approved by the Clinical Science and Quality Committee (CSQC).±
Composite Intent: An Advisory Group member asked what the developer's intent was when creating the composite.	They used the composite methodology with the goal of facilitating public reporting and enhancing its utility for patients/stakeholders. CMS has previously promoted the use of composite measures to ease the burden of comparing hospital quality for patients. ±
Individual Component Reporting: An Advisory Group member asked whether the measure is reported as one composite score, or if scores are also reported for the individual components.	As each sub-component is its own measure, the individual components are also reported.
Component Testing: An Advisory Group member asked if the developer had testing information for each component.	They provided testing data as required for the composite model. Additional testing information is available for the individual measures (mortality [CBE#0133], bleeding [CBE #2459], inpatient cardiac rehabilitation [CBE# 0642, CBE #0643], discharge on guideline directed medical therapy [CBE #0964]) that already have undergone CBE endorsement. Additional information is also included for the acute kidney injury (AKI) component measure in the supplemental attachment in the measure application.±
Combined Variables: An Advisory Group member asked why non-ST-elevation myocardial infarction (NSTEMI) and unstable angina (UA) are combined.	In alignment with the 2014 clinical practice guidelines (CPGs), "At presentation, patients with UA and NSTEMI can be indistinguishable and are therefore considered together in this CPG." ±
Importance and Patient-Centered Care: A few Advisory Group members said this is an important topic. One Advisory Group member highlighted that the measure may help inform care, allowing people to make decisions and actively participate in their treatment.	Not applicable.
Additional Component: An Advisory Group member suggested including an element that captures referral, patient education, and patient preference.	The developer did not respond to this comment.

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

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

Discussion Category	Key Themes	Source of Comment	Summary of Comments
Dissenting	Weighting	Advisory Group	The Advisory Group questioned whether the weighting of the individual components is appropriate.
	Feasibility	Staff Assessment	This domain is rated as "Not Met but Addressable" due to incomplete discussion of the feasibility assessment performed, adjustments made to the specifications based on feasibility, and requirements for participating in the registry to report the measure.
	Reliability	Staff Assessment	The current reliability metrics do not meet the established thresholds, indicating potential issues with the consistency and accuracy of the results across different settings and populations. However, the identified limitation is deemed addressable, as the developer may consider identifying sources of prior evidence for the additional 16 data elements.
Mixed	Importance	Advisory Group; Staff Assessment	A few Advisory Group members said this is an important topic and may help inform care, allowing people to make decisions and actively participate in their treatment. The staff assessment rated this domain as "Not Met but Addressable" due to a sparse logic model and potentially insufficient patient input.

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 #4650: Full Responses Written by the Developer

Feedback/Questions	Full Developer Response
Reliability Table: An Advisory Group member asked if the developers were saying that the smallest facilities do the best job and the largest facilities do the worst? Table 2 under Reliability shows the total volume goes up across the 10 deciles from 91,207 to 888,133. However, performance appears to worsen from 7.8% (smallest facilities) to over 45% (largest facilities) across these volume deciles. This finding goes against a lot of evidence from other settings that more experienced, higher-volume facilities do better on most quality measures.	After additional review, we discovered that the performance scores in Table 2 were inadvertently copied from Table 1. The IUR values in Table 2 as submitted are correct, and a revised Table 2 is given below (Table A1), with the correct performance scores for each decile. As queried in the comment, there is a small improvement in performance score noted from decile 1 (small facilities) to decile 10 (large facilities). Of note, in the MIN column, 37 facilities (with a total of 2,282 patient months) have the minimum of 11 patients per facility to be included in the measure.

Table A1: Accountable Entity–Level Reliability Testing Results by Denominator-Target Population Size

	Overall	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max
Reliability (IUR)	0.76723	0.35573	0.47582	0.59048	0.65226	0.69369	0.72799	0.75707	0.78269	0.81017	0.83858	0.8833	0.9512
Mean Performance Score (deciles of facility size)	23.01%	27.02%	25.17%	23.04%	23.53%	23.19%	22.46%	23.01%	22.59%	22.76%	22.40%	22.07%	33.57%
N of Facilites	7497	37	745	775	739	696	795	723	763	754	757	750	1
N of Reporting Patient-Months	3,758,302	2,282	91,207	159,305	203,405	233,587	318,102	338,400	418,557	495,687	611,919	888,133	2,237

CBE #4595: Full Responses Written by the Developer

Feedback/Questions	Full Developer Response
Unintended Consequences: An Advisory Group member asked the developer to speak on potential unintended consequences and mitigation strategies.	As noted in our submission, we did not identify any unintended consequences during measure development or model testing. However, we are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other potential unintended consequences for patients. There has been increased focus on the timeliness of care provided to stroke patients, and hospitals can opt for one of four different levels of stroke center certification (each with its own set of process and outcome measures) as assessed by The Joint Commission (TJC). All but two measures in TJC's certification program are process measures; there are two outcome measures based on the Modified Rankin Score (mRS) at 90 days; however, those outcome measures can, by definition, only apply to patients who survived the stroke. There are also process measures associated with quality improvement efforts around stroke (e.g., Get With the Guidelines). Importantly, none of these processes or outcome measures are reported to the public at the facility level. We note that it is possible for a hospital to perform well on mRS measures but have poor performance on the Stroke Mortality measure. Other than the Stroke Mortality measure currently under review in this submission (and the prior Fee-For-Service [FFS]-only version, which is currently publicly reported), no other stroke mortality measures are publicly reported. This Stroke Mortality measure will be publicly reported to patients and other stakeholders on *Care Compare* (and currently, the prior version of this measure, with only FFS patients, is publicly reported). This illustrates the measurement gap that is filled by this publicly reported, risk-standardized mortality measure for patients with ischemic stroke as it complements the other stroke-related measures in the quality landscape and can be used by those organizations to track progress on quality improvement efforts. It also serves as a public reporting and transparency/accountability measure for public use.

Feedback/Questions	Full Developer Response
Unintended Consequences on Rural/Under-Resourced Communities: An Advisory Group member asked if this measure negatively impacts facilities that are not stroke centers and serve rural or under-resourced communities.	We believe that it is important for all hospitals to be able to track their performance on the outcome of stroke mortality. However, hospitals that admit fewer than 25 stroke patients during the performance period do not have their performance scores publicly reported. They do, however, receive, confidentially, their measure score, in addition to patient-level information about the stroke patients that were admitted to their hospital, to support quality improvement efforts. We note that if a patient with a stroke enters an emergency department and is transferred to another hospital, the stroke outcome is attributed to the hospital that admits the patient (the receiving hospital in this case). If a patient is admitted to a hospital and subsequently transferred to a second hospital, the mortality outcome is attributed to the first hospital.
Underlying Risk: An Advisory Group member asked what is responsible for the underlying risk of stroke for patients in the cohort.	Clinicians and stakeholders, including the American Heart Association, American Stroke Association, and other professional organizations, highlight the importance of including an assessment of stroke severity in risk-adjustment models of stroke mortality. Several studies have demonstrated that initial stroke severity is the strongest predictor of mortality in ischemic stroke patients. Furthermore, testing from development as well as more recent testing demonstrate that adjusting for stroke severity using the NIH Stroke Scale (NIHSS) from administrative claims improves discrimination of the stroke mortality risk model. The NIHSS, which was created in 1989, is widely used in routine stroke care. Beginning in October 2016, NIH Stroke Scale score ICD-10-CM codes became available in administrative claims (as secondary diagnoses). CMS added the NIHSS to the Stroke Mortality measure with the 2022 public reporting of the measure on Care Compare. In the respecified measure, variables for the stroke mortality measure were selected using an empiric approach described in the “Risk Adjustment” section of the CBE submission. Following the empiric selection of risk variables, we added the NIHSS (see “All Figures and Tables Stroke Mortality” attachment); the NIHSS is one of more than 95 other variables in the final risk model. The NIHSS ranges from 0-42; we use the numerical NIHSS score in the regression model. For every unit increase of NIHSS score the odds of death increase by 4%.

Feedback/Questions	Full Developer Response
	Risk variables (ICD-10 codes) with the highest odds ratios include, during the index admission: Secondary malignant neoplasm of liver and intrahepatic bile duct (OR of 6.51); Compression of brain (OR of 2.51); NonST elevation (NSTEMI) myocardial infarction (OR of 2.03); Acute respiratory failure, unspecified whether with hypoxia or hypercapnia (OR of 3.14); Acute respiratory failure with hypoxia (OR of 2.26), and encounter for palliative care (OR of 22.81). Please see Table 9 in the "All Figures and Tables Stroke Mortality" attachment for more details.

CBE #0753: Full Responses Written by the Developer

Feedback/Questions	Full Developer Response
Unresolved Issues: An Advisory Group member asked why surgical attire is considered an unresolved issue. A patient participant Advisory Group member shared a personal experience where a new protocol was being implemented during surgery. She noted that unresolved issues could undermine trust and safety in the health care setting.	Please see question 2.2 in the measure submission regarding the evidence to support the measure. Literature is provided to link structure and processes to the desired health outcome of SSI prevention. The SHEA/IDSA/APIC Strategies to prevent surgical site infections in acute-care hospitals: 2022 Update provides a summary of recommendations that acute care hospitals may implement to prevent SSIs. The "Recommendations are categorized as either (1) essential practices that should be adopted by all acute-care hospitals or (2) additional approaches that can be considered when hospitals have successfully implemented essential practices and seek to further improve outcomes in specific locations and/or patient populations. Essential practices include recommendations in which the potential to affect HAI risk clearly outweighs the potential for undesirable effects. Additional approaches include recommendations in which the intervention is likely to reduce HAI risk but there is concern about the risks for undesirable outcomes, recommendations for which the quality of evidence is low, or recommendations where the evidence supports the effect of the intervention in select settings (e.g., during outbreaks) or for select patient populations." The document also includes a list of unresolved issues to potentially reduce SSI, including use of surgical attire. The guideline provides the following explanation for this unresolved issue: "Although there are longstanding traditions and opinions regarding surgical attire in the operating room, no strong evidence exists for many of them. It has not been demonstrated that surgical attire affects SSI rates. One approach to managing issues pertaining to surgical attire is to form a multidisciplinary body including infection control, surgery, nursing, and anesthesia to discuss and agree to some sensible, not overly aggressive or cumbersome attire standards, and to establish policies and procedures that are compliant with state and CMS requirements." [Calderwood et al. 2023, page 29] Reference: Calderwood MS, Anderson DJ, Bratzler DW, Dellinger EP, Garcia-Houchins S, Maragakis LL, Nyquist AC, Perkins KM, Preas MA, Saiman L, Schaffzin JK, Schweizer M, Yokoe DS, Kaye KS. Strategies to prevent surgical site infections in acute-care hospitals:

Feedback/Questions	Full Developer Response
	2022 Update. Infect Control Hosp Epidemiol. 2023 May;44(5):695-720.
Risk-Adjustment Model: The Advisory Group talked about several facets of risk adjustment, including: • **Facility-Level Unintended Consequences:** A few Advisory Group members were concerned that the measure would unfairly penalize certain types of facilities. One Advisory Group member said they believed the measure would reward large non-trauma hospitals and penalize smaller hospitals that are in rural or under-resourced areas because the measure uses hospital/facility-level characteristics in the risk model. They noted that it was not appropriate to use hospital as a fixed effect in the risk-adjustment model. • **Patient-Level Risk Adjustment:** Several Advisory Group members said they believed that the measure should be risk adjusted by patient-level characteristics. An Advisory Group member pointed out that the if the measure was risk adjusted by patient-level characteristics and did not use hospital as a fixed effect, the measure would not reward large hospitals. • **Hierarchical Clustering:** An Advisory Group member stated they believed the model should have hierarchical clustering. • **Feature-Selection Methods:** An Advisory Group member stated that the feature selection methodology used in risk adjustment of this measure is outdated and biased. They suggested manuscripts (2019 in The American Statistician and 2020 in the International Journal of Epidemiology) on stepwise regression creating unfair risk-adjustment models for the Recommendation Group. • **Moderate C Statistic:** While a few Advisory Group members criticized the model for having a moderate C statistic of 0.6; a patient participant pointed out that while the number is smaller, the committee at least knows they are dealing with truthful numbers and it something that can be improved upon. Overall, a few Advisory Group members wished they had been given more methodology information by the developer in the submission materials.	**On Using a Fixed-Effect Model:** The SSI risk models following each colon and hysterectomy procedures are currently constructed as fixed-effect logistic regression models. We have used fixed versus mixed-effect (hierarchical) models for several reasons: First, the observed-to-predicted construction of each measure requires a predictive model be statically used for tracking changes over time. This feature is especially important for accurately tracking progress improvement at the facility, state, and federal levels. The NHSN has a long-standing practice for being used to measure performance over time. Second, to include any random effect terms would impose facility adjustment that assumes no changes in the composition of reporting facilities for measurement in years beyond the baseline year. In other words, a model that includes a random effect would not generalize to future years of reported data. This adjustment would impose a unique annual recalibration for each facility and year that would severely impact assessing temporal changes. Third, new facilities enrolling and reporting these data to NHSN would neither be able to have their predicted SSIs calculated nor report their SIR if a mixed model was used. Last, the NHSN surveillance system provides end users with immediate calculation of these SIR measures as well publishing the model parameters and instructions to create their own analyses, including any individual predicted probabilities at the patient-procedure level. Since these are very straightforward with a fixed effect model, it allows the users to simply sum observed and predicted SSIs for easy calculation and comprehension of the SIR. A mixed model eliminates this ability and avoids real-time calculation and comparisons at the patient and facility level. **On the C Statistic:** The C-statistics for both SSI models were 0.635 for colon and 0.624 for hysterectomy procedures, respectively. To improve the C-statistics beyond this level requires a much heavier data collection burden of factors that among them includes a few additional factors significantly associated with SSI, and yet, may yield marginal increases.

Feedback/Questions	Full Developer Response
	Furthermore, many measure models may have C-statistics in this range that are or have been used for performance measurement. Examples include: 30-day readmission for pneumonia (c=0.63, J Hosp Med. 2011 Mar; 6(3): 142-50), Heart Failure (c=0.61, Circ Cardiovasc Qual Outcomes. 2009 Sep; 2(5): 407-13) and AMI Failure (c=0.64, Circ Cardiovasc Qual Outcomes. 2009 Sep; 2(5): 407-13). Additionally, smaller cohort studies of SSI after abdominal surgery yielded only slightly higher C-statistics of 0.72 (J Surg Res. 2017 Sep;2017: 153-159). CDC/NHSN staff are seeking to implement several patient-level measures collected in a manner that will not significantly increase manual data collection burden.
Trauma Cases: An Advisory Group member asked why the measure captures trauma cases and whether trauma is considered in risk adjustment.	Facility factors such as bed size were allowed into the model-building process to achieve improvement in the SSI prediction models. These potentially included facility factors help complement the list of patient procedure-level factors and can serve as surrogates for patient risk not captured by the existing patient procedure factors. In our colon model, we found that both larger hospitals and those designated as having a major medical school affiliation had marginally higher SSI risk, and this difference persisted after accounting for significant patient procedure factors although with a smaller impact on predicted risk. To observe higher SSI risk in a larger or major medical school affiliated hospital likely indicates they are serving a higher acuity patient mix. The process for determining cut points is data driven and involves potentially combining levels of SSI risk across ordinal deciles. It is grounded in a uniformly applied framework and avoids arbitrary decisions while allowing for thoughtful consideration from NHSN statisticians. In this specific case, differences in SSI risk were significantly different at greater than or equal to 319 beds, which is a numerical value to separate combined deciles. The NHSN Patient Safety Component SSI Protocol is available to all acute-care/critical-access hospitals, including those performing surgeries related to trauma, which are relatively low volume [approximately 3.2% with the 2022 rebaseline]. The 2022 baseline for the complex 30-day model uses trauma, among other factors, as a predictor of infection following colon surgery (COLO). COLOs reported

Feedback/Questions	Full Developer Response
	as Trauma = Yes are compared to those reported with Trauma = No to assess the likelihood/risk of infection. The results of the 2022 baseline model for COLO using the complex 30-day SSI SIR model show that COLOs reported as Trauma = Yes is the risk group while those reported with Trauma = NO is the referent group. The factor trauma is an important risk indicator for SSIs in COLO surgical procedures, as it was found significant ($p<0.001$) in our multivariable model. Trauma patients therefore are given additional patient risk in calculating the SIR. The summation of patient risk is calculated in the denominator of the SIR (expected occurrence of SSIs). Therefore, a HIGHER patient risk will bring down the SIR ratio. While it may not be possible to control for the trauma itself, surveillance for these events is a critical part of developing effective strategies for prevention of SSIs. Details of the 2022 rebaseline Complex 30-day model for COLO and HYST are published in the new SIR Guide linked here.

CBE #4580: Full Responses Written by the Developer

Feedback/Questions	Full Developer Response
Weighting: A few Advisory Group members expressed concern or asked for more information about how the components are weighted within the composite measure. One Advisory Group member felt that the components should be weighted more equitably, given that they all reflect processes that have been shown to improve patient outcomes.	The goal of the weighting process was to achieve a clinically meaningful balance between the assigned weights of process and outcome metrics, both of which comprise the CathPCI Quality Composite. The outcome metrics, mortality, AKI, and bleeding, are weighted at 35%, 20%, and 20%, respectively. Process metric 45 (cardiac rehab) is weighted at 5%. Finally, the process metrics 4 (PCI within 90 minutes) and 38 (GDMT at discharge) are weighted at 10% each. The ACC's Metrics and Reporting Methodology (MRM) sub-committee and data analytical center considered multiple options in weighting the components of the PCI Quality of Care composite measure. In addition, they consulted with the scientific lead of Yale CORE's CMS star rating team. MRM agreed that outcome measure components should have more weight than process measures. This agreement was based on their understanding of the patient perspective that procedural outcomes are more important than the process taken to achieve the final outcome. Five unique weighting options were proposed and reviewed (Table A2). The p-scores increase slightly as more weight was placed on outcomes (specifically the mortality measure). MRM voted and selected approach #3. The descriptive statistics associated with each scenario are listed below (Table A3). Within scenario 3, the outcomes measures combined account for 75% of the composite weight with process measures accounting for 25%. The highest weight was placed on risk of in-hospital mortality (35%).
Composite Intent: An Advisory Group member asked what the developer's intent was when creating the composite.	ACC understands this comment as asking for clarity on the methodologic question of choosing a composite measure over reporting each component measure separately. We are answering with this perception in mind. Based on the 2006 IOM (Institute of Medicine; now National Academy of Medicine) report on performance measurement, Performance Measurement: Accelerating Improvement, it was noted that composite

Feedback/Questions	Full Developer Response
	measures can enhance measurement to extend beyond tracking performance on separate measures and can provide a potentially deeper view of the reliability of the care system. (Nolan et al., 2006). The composite measure design was in response to this report and in alignment with the National Quality Forum's (NQF) Consensus Report "Composite Measure Evaluation Framework and National Voluntary Consensus Standards for Mortality and Safety—Composite Measures" (Krumholz et al., 2009). In summary, CMS promoted the use of composite measures and as discussed in the original application intent section, one score was identified as having better utility for patients than comparing multiple scores. Nolan T, Berwick DM, All-or-none measurement raises the bar on performance, JAMA, 2006;295(10):1168-1170. Krumholz, H., Fowles, JB. Amundson, G., et al., 2009. Composite Measure Evaluation Framework and National Voluntary Consensus Standards for Mortality and Safety—Composite Measures. Washington DC: National Quality Forum.
Component Testing: An Advisory Group member asked if the developer had testing information for each component?	Testing information on the component measures is not included within this application. Component measure testing data was not requested within this application. In composite measure-specific applications, Battelle and NQF had previously instructed measure developers to "Provide performance scores on the measure as specified." ACC followed this guidance in providing testing data for Measure 4580 as specified as a composite measure. However, four of the component measures already are CBE endorsed (mortality, bleeding, inpatient cardiac rehabilitation, discharge medication). The other two, PCI in 90 minutes and acute kidney injury measures, are not CBE endorsed. For additional information, the AKI manuscript is included in the original application attachment.
Combined Variables: An Advisory Group member asked why non-ST-elevation myocardial infarction (NSTEMI) and unstable angina (UA) are combined.	The 2014 NSTE-ACS clinical practice guidelines (CPG) is a full revision of the 2007 ACCF/AHA CPG for the management of patients with unstable angina (UA) and non–ST-elevation myocardial infarction (NSTEMI) and the 2012 focused update. The new title, "Non–ST-Elevation Acute Coronary Syndromes," emphasizes the continuum

Feedback/Questions	Full Developer Response
	between UA and NSTEMI. At presentation, patients with UA and NSTEMI can be indistinguishable and are therefore considered together in this CPG. Amsterdam EA, Wenger NK, Brindis RG, Casey DE Jr, Ganiats TG, Holmes DR Jr, Jaffe AS, Jneid H, Kelly RF, Kontos MC, Levine GN, Liebson PR, Mukherjee D, Peterson ED, Sabatine MS, Smalling RW, Zieman SJ. 2014 AHA/ACC guideline for the management of patients with non–ST-elevation acute coronary syndromes: a report of the American College of Cardiology/ American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2014;64:e139–228.

Table A2. Scenarios proposed for weighting component measures

			Scenario options				
Metric	**Description**	**Type**	**1**	**2**	**3**	**4**	**5**
Metric 1	In-hospital mortality	Outcome	30%	35%	35%	45%	55%
Metric 4	PCI w/in 90 min	Process	10%	5%	10%	5%	5%
Metric 38	GDMT at DC	Process	10%	5%	10%	5%	5%
Metric 39	Acute Kidney Injury	Outcome	20%	25%	20%	20%	15%
Metric 40	Bleeding	Outcome	20%	25%	20%	20%	15%
Metric 45	Cardiac Rehab	Process	10%	5%	5%	5%	5%

Table A3: Descriptive statistics for various weighting scenarios

	1	2	3 (selected)	4	5
Mean	85.42	84.90	86.64	86.73	88.57
STDV	+/- 5.28	+/- 4.91	+/- 4.37	+/- 4.28	+/- 3.74
IQ1- IQ3	82.11-89.34	82.06-88.33	84.21-89.66	84.32-89.76	86.37-91.23
Median	86.22	85.40	87.04	87.16	88.95

Within scenario 3, the outcomes measures combined account for 75% of the composite weight with process measures accounting for 25%. The highest weight was placed on risk of in-hospital mortality (35%).

Figure A1: Distribution of various weighting scenarios

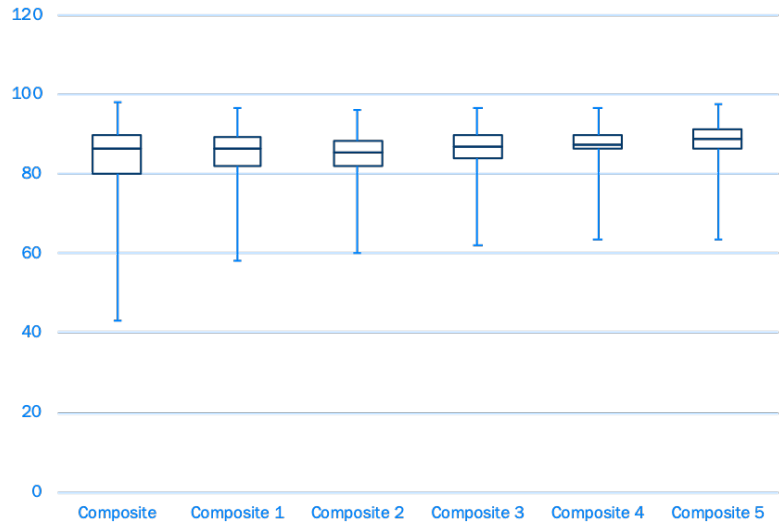


Figure A2: Distribution of measures