Physician Cost Measures and Patient Relationship Codes (PCMP) Technical Expert Panel

August 12, 2025

Summary Report

October 2025

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

TABLE OF CONTENTS

1 INTRODUCTION

The Centers for Medicare & Medicaid Services (CMS) contracted Acumen, LLC ("Acumen") to develop and maintain cost measures and Patient Relationship Category codes (PRCs). The contract name is Physician Cost Measures and Patient Relationship Codes (PCMP). The contract number is 75FCMC18D0015, Task Order 75FCMC24F0142. As part of this work, we convene a standing Technical Expert Panel (TEP) to provide input on overarching issues across all activities. This report summarizes the TEP meeting on August 12, 2025. Section 1 outlines the structure and composition of the panel. Section 2 summarizes each session's presentation, TEP member input, and key takeaways. Finally, Section 3 outlines the next steps for this project.

1.1 Project Summary

Under the PCMP contract, Acumen develops episode-based cost measures for potential use in the Merit-based Incentive Payment System (MIPS) to meet the requirements of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA), and maintains the population-based measures in use in MIPS (Total Per Capita Cost [TPCC] and Medicare Spending Per Beneficiary [MSPB] Clinician). Acumen also supports implementation of Patient Relationship Codes (PRCs), which were previously developed in alignment with requirements established in MACRA. Additionally, Acumen maintains the Medicare Spending Per Beneficiary (MSPB) Hospital measure used in the Hospital Value-Based Purchasing (VBP) program and the Hospital Inpatient Quality Reporting (Hospital IQR) program.

1.2 Standing TEP Member and Meeting Information

The PCMP TEP comprises 23 members with a range of perspectives and areas of expertise, including experts in health care, payment policy, payment models, and performance measurement; clinicians across many specialties; and patient advisors who share their perspectives from lived experiences. Please see Appendix A for the TEP member composition list.

This TEP met for the first time on December 18, 2024 to discuss cost measure development prioritization; developing value measures to align with cost measures; and testing and next steps for PRCs.[1] The TEP met again virtually on August 12, 2025 from 10:00 am to 1:00 pm ET, with 17 of the 23 members in attendance. The TEP meeting focused on cost measure evaluation. The first session provided a recap of PCMP activities since the previous

[1] Centers for Medicare & Medicaid Services. "Physician Cost Measures and Patient Relationship Codes (PCMP) Technical Expert Panel December 18, 2024 Summary Report." August 2025. https://mmshub.cms.gov/sites/default/files/2025-05-08-pcmp-tep-summary.pdf.

meeting. The second session provided an overview of measure evaluation criteria used in different contexts during the measure lifecycle. The third session discussed how the validity criterion applies to cost measures. The fourth session discussed balancing the various measure evaluation criteria. After the meeting, Acumen followed up with TEP members to gather their written feedback and recommendations through a post-meeting survey.

2 DISCUSSION SUMMARY

This section summarizes Acumen's presentation, TEP member discussions, TEP member post-meeting survey responses, and key takeaways for each of the TEP meeting sessions. The discussion summaries consolidate related feedback but do not necessarily represent consensus. Section 2.1 recaps the presentation about measure evaluation criteria. Section 2.2 summarizes the session regarding validity evaluation criterion. Section 2.3 summarizes the session about balancing the various measure evaluation criteria.

2.1 Overview of Measure Evaluation Criteria

Acumen provided an overview of CMS materials and processes related to measure evaluation criteria and reviewed guidance on meeting measure evaluation criteria. Section 2.1.1 summarizes Acumen's presentation. Since the purpose of this session was to provide members with necessary context, TEP members discussed these criteria during the subsequent sessions.

2.1.1 Summary of Presentation

Acumen provided an overview of measure evaluation criteria used by various CMS and Consensus-Based Entity (CBE)[2] processes and sources throughout the measure lifecycle:

- The CMS Measures Management System (MMS),
- The Measures Under Consideration (MUC) process,
- The Partnership for Quality Measurement (PQM) Endorsement & Maintenance (E&M) process, and
- The PQM Pre-Rulemaking Measure Review (PRMR) process.

Measure evaluation criteria set standards for how to assess a measure. Acumen outlined how these measure evaluation criteria are applied to cost measures throughout the measure lifecycle. At the outset of development, Acumen and CMS consider evaluation criteria to prioritize measure concepts that have potential to meet all the evaluation criteria. Throughout the measure development specification process, Acumen iteratively reviews and tests the measure

[2] The Medicare Improvements for Patients and Providers Act of 2008 requires the Department of Health and Human Services (HHS) to contract with a consensus-based entity regarding performance measurement. Battelle currently serves as the CMS CBE and uses its Partnership for Quality Measurement (PQM) to incorporate feedback from interested parties to evaluate and endorse quality and cost measures.

against the evaluation criteria. Once measure development is complete, the CBE evaluates the measure against evaluation criteria as part of the PRMR and E&M processes.

Acumen noted similarities and differences in evaluation criteria across these sources and processes. All aforementioned sources and processes include importance, feasibility, and usability criteria. However, there are notable differences between the definitions of each criterion depending on the source and in some cases, there are additional evaluation criteria included.

For example, the MMS guidance includes four key categories of evaluation criteria: importance, feasibility, usability and use, and scientific acceptability. At a high level, the PQM E&M process criteria are similar to the CMS MMS guidance, but definitions of each criterion vary. The PQM E&M process's scientific acceptability criterion mentions assessing reliability and validity of the measure, while the CMS MMS guidance on scientific acceptability assesses reliability, validity, exclusions, and risk adjustment. The PQM PRMR process maintains importance, feasibility, and usability criteria under a broader "meaningfulness" category. However, instead of a scientific acceptability criterion, PRMR has a validity/reliability criterion and has added criteria for conformance, appropriateness of scale, and time to value realization that are not present in other evaluation guidance.

Finally, Acumen highlighted differences in how these evaluation criteria are applied, particularly that only the reliability criterion has a numeric threshold. All other criteria utilize qualitative guidance. For example, guidance for the importance criterion asks evaluators to assess how evidence-based or crucial a measure is, or how important it is to health care quality or cost. As a result, there is some interpretation required from both measure developers and the measure evaluators to determine if the criterion has been met. For reliability, the MIPS program requires cost measures being added to the program to have a mean reliability of 0.4 or greater. For the purposes of MIPS, CMS outlined that a mean reliability of 0.4 to 0.7 is considered moderate reliability, while a reliability greater than 0.7 is considered high reliability.[3] In contrast, the PQM E&M process requires 70% or more of entities to have either a signal to noise/inter-unit reliability greater than or equal to 0.6, or a split-half reliability (intraclass correlation [ICC]) of greater than or equal to 0.6. In a December 2024 public posting, PQM also listed 0.6 as their numeric threshold for reliability for the PRMR process.[4] However, recent publications, including

[3] CY 2017 QPP Final Rule (81 FR 77169 through 77171). https://www.federalregister.gov/d/2016-25240/p-2170.
[4] Partnership for Quality Measurement. "Pre-Rulemaking Measure Review (PRMR): Preliminary Assessment Process and Content Overview." December 2024. https://p4qm.org/sites/default/files/2024-12/PRMR-Overview-PA-Methodology.pdf.

the July 2025 Guidebook of Policies and Procedures for PRMR and Measure Set Review (MSR), do not mention a specific numeric threshold.[5]

2.2 Validity and Cost Measures

This session focused on reviewing methods to evaluate, demonstrate and communicate validity to interested parties. Section 2.2.1 recaps Acumen's presentation, Section 2.2.2 recaps the TEP members' discussion, Section 2.2.3 summarizes the TEP members' post-meeting survey responses, and Section 2.2.4 contains key takeaways.

2.2.1 Summary of Presentation

Generally, cost measures are considered valid when they assess variation in cost performance within the clinician's reasonable influence, and measure what they intend to measure. Acumen summarized various ways that validity criterion is defined, and what evidence is required to establish it, across activities throughout the measure lifecycle. The CMS MMS states that the measure developer must conduct validity testing at the data element-level and measure score-level and not solely rely on face validity testing, unless the measure is new (i.e., not currently used in a CMS program). MMS also provides information on the various types of empirical validity testing that measure developers can use, such as construct, criterion, or discriminant validity.

While MMS provides overarching guidance on validity, PQM defines validity as "an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy of appropriate interpretations and actions on the basis of [the measure]." [6] Documentation supporting the MUC submission process refers to the MMS guidance and definition, which states that "the term 'validity' has a specific application known as test validity, which refers to the degree to which evidence, clinical judgment, and theory support interpretations of a measure score."[7] While the two definitions align, the PRMR and E&M processes request different evidence. For example, the PQM PRMR process requests evidence that there are ways to improve the measure focus. In contrast, the E&M process requests evidence of a valid analytic approach and that threats to validity are not present. Acumen must reconcile differences in evidence requested or evaluation criteria to develop effective cost measures.

[5] Partnership for Quality Measurement. "Guidebook of Policies and Procedures for Pre-Rulemaking Measure Review (PRMR) and Measure Set Review (MSR)." July 2025. https://p4qm.org/sites/default/files/2025-07/OP2-PRMR-MSR-Final-Multi-Stakeholder-Group-Guidebook-of-Policies-and-Procedures-508_1.pdf.
[6] Partnership for Quality Measurement. "Guidebook of Policies and Procedures for Pre-Rulemaking Measure Review (PRMR) and Measure Set Review (MSR)." July 2025. https://p4qm.org/media/1331.
[7] CMS Measures Management System. "Validity." July 2025. https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/scientific-acceptability/validity.

Acumen stated that the purpose of cost measures influences developer approaches to validity and detailed several examples of non-empirical, or theoretical, and empirical approaches we employ to demonstrate validity:

- Theoretical approaches mentioned were environmental scans and literature reviews; input from experts and persons with lived experience; and logic and conceptual models.
- Empirical approaches discussed were face validity votes; cost driver analyses; mediation analyses; measure correlations; and risk adjustment analyses.

Acumen reviewed the approach examples, emphasizing benefits and drawbacks and detailing different use cases. A literature review may be useful in assessing the importance of a cost measure, evaluating the prevalence of a condition. Additionally, Acumen could seek clinical expert input or conduct empirical analyses to understand variation in cost performance within reasonable influence of the clinician. The best approach to demonstrate validity may be a combination of different theoretical and empirical validity assessment methods.

Acumen presented the following questions for discussion:

- What is a "valid" cost measure?
- How should we evaluate or demonstrate validity of cost measure, versus the applications of cost measure?
- How should we consider the various empirical and theoretical assessments?
- For example, should one be weighted more or less heavily than another?
- What information is helpful to show that threats to validity have been mitigated?
- Are there additional steps we should consider during development to help ensure validity of specifications? Additional testing methods?
- How should we interpret differing opinions between clinician expert workgroup members, review panels, and other interested parties?

2.2.2 TEP Member Discussion

TEP members discussed different strategies to ensure cost measure validity, including risk adjustment and cost and quality alignment. They also recommended steps that Acumen can take, including additional testing and enhanced communication efforts. Overall, members suggested that cost measures are more valid when they assess variation in cost performance within the clinician's reasonable influence and measure what they intend to measure.

TEP members particularly emphasized the importance of an appropriate risk adjustment model to a measure's validity. Risk adjustment ensures that the measure assesses costs that are within a clinician's reasonable influence. TEP members stated that risk adjustment is key in ensuring validity, so that measures account for potential increased costs for higher-risk patients and avoid potential unintended consequences of care stinting for high-risk patients. Other

members mentioned the importance of adjusting for condition severity in cost measures. One member stated that this can be challenging for measures, depending on the available coding options. Condition severity is better able to be accounted for through risk adjustment when specific International Classification of Diseases, 10th Revision (ICD-10) diagnosis codes exist for severity, compared to conditions where ICD-10 diagnosis codes do not exist for severity. TEP members also suggested adjusting for geographic impacts in cost measures due to differences in medical malpractice laws and other differences outside of the control of the clinician.

Several TEP members also discussed cost and quality alignment as a facet of measure validity. One member highlighted that cost measures incorporate certain quality metrics or concepts within its design. They recommended that this be considered more specifically during cost measure development, as intentionally capturing these metrics can help mitigate criticisms of cost measures that they are not drivers of improved care and will do more for the acceptability of this type of measurement. This member suggested that cost measures can help improve quality by encouraging clinicians to alter their behavior in a positive way, improving quality while lowering cost. They also recommended assessing the correlations between cost and quality measure to demonstrate this relationship. One member suggested that guideline-directed care is not always the most cost-effective. They recommended accounting for the use or frequency of these necessary services (e.g., drugs, therapeutic interventions) captured by cost measures to discourage care stinting.

Additionally, TEP members discussed testing that would be helpful to confirm a measure's validity. A few members suggested adjusting for Medicare and Medicaid dual eligibility and other non-clinical patient factors in cost measures.

Several TEP members discussed the effects of cost measures on small group practices, which may not currently be captured due to case minimum requirements. Several members emphasized the need to include more small group practices with lower volume providers in cost measure attribution, and that this may enhance cost measure validity if attribution more accurately reflects the clinician population. Members suggested that Acumen analyze measure performance and stratify by large or small clinician groups to see if they are performing differently, and consider risk adjusting to account for any differences observed. One member also suggested sub-grouping group practices under the same Taxpayer Identification Number (TIN) by specialty, so that cost measures do not assess a group of clinicians with different specialties that may be in a group practice together as one unit, when care patterns may differ.

Finally,, TEP members provided feedback on demonstrating validity for cost measures. Several panelists emphasized the need for better communication about cost measures with the clinician community. They highlighted that a measure must be well defined to asses validity and that the clinician community is not aware of the amount of testing and iterative development

involved in cost measure specifications,, particularly for the risk adjustment methodology. TEP members emphasized how that the thoroughness and rigor of risk adjustment models should be better communicated to clinicians. Other members recommended that Acumen provide information on validity to clinicians, similar to the TEP presentation, and collaborate with specialty societies to disseminate information about cost measures and development.

2.2.3 TEP Member Survey Responses

In the post-meeting survey, TEP members discussed the characteristics of a valid cost measure. Members felt that a valid cost measure should be comprehensive, accurate, reliable, and credible, reflecting costs that are clinically meaningful, fairly attributable, and risk-adjusted to account for patient heterogeneity and factors beyond a clinician's control. They recommended the measures focus on costs under the reasonable influence of providers, ensuring transparency and fairness to avoid penalizing clinicians for factors outside of their control. Additionally, TEP members noted that valid cost measures should be actionable, aligned with quality outcomes, and promote value-based care that improves health outcomes without unfair cost shifting. Lastly, valid cost measures should be feasible to implement, reproducible, and able to accurately distinguish between high and low resource use over an appropriate timeframe, enabling clinicians to make informed decisions that reduce unnecessary costs while maintaining or enhancing care quality.

In terms of the theoretical and empirical approaches to validity assessment that were discussed during the TEP meeting, members ranked the top approaches based on their effectiveness at demonstrating cost measure validity to interested parties. The top validity approaches chosen by TEP members based on a ranked choice voting system were (i) Logic and Conceptual Models, (ii) Input from Experts and Persons with Lived Experience, and (iii) Cost Driver Analyses. Respondents highlighted several additional theoretical and empirical approaches that could enhance the validity assessment of cost measures beyond those discussed during the TEP meeting. These approaches include ongoing evaluation of the predictive value of risk adjustment models to build clinician trust, the use of multi-trait multi-method (MTMM) analysis, logic model validation, known-groups testing, replication studies in diverse settings, and simulation studies to strengthen robustness. Some members also suggested incorporating external reviews and endorsements from entities such as specialty societies as part of the validity process. One member recommended examining frameworks used by international organizations, such as the National Institute for Health and Care Excellence (NICE), for further insights.

To help ensure the validity of cost measure specifications, TEP members recommended additional steps to consider during development. One member reiterated the importance of aligning cost and quality. One member recommended comparing cost stratified by zip code to confirm the measure accounts for environmental factors affecting costs. Testing for changes over

time post-implementation and incorporating logic models, stress-testing, cross-database replication, and stakeholder usability assessments were also recommended. Testing methods like sensitivity analyses, split-sample and temporal validation, and simulation studies could further strengthen validity. Members emphasized addressing the limitations of the common one-year measurement period and prioritizing alignment between cost and quality measures, ensuring both are reliable, valid, feasible, and focused on the same patient populations. Additionally, members felt that comprehensive risk adjustment was important to ensure fairness and accuracy.

To demonstrate that threats to validity have been effectively mitigated, TEP members emphasized the importance of transparency in measure development, rigorous specification, and robust risk adjustment. Key helpful information includes detailed cost driver and risk adjustment analyses, documentation of exclusions, data consistency checks, handling of missing data, and validation of attribution. Assessments of population homogeneity, comparisons between variations in cost and quality outcomes, and analysis of physician groupings by patient risk further support validity. Members felt that providing frequent updates to expert panels and public disclosure of the validation process were also essential. Additionally, extending measurement windows for chronic conditions, analyzing data across multiple years and sources, and monitoring cost trends across conditions help address temporal and contextual factors. Finally, sensitivity analyses, independent replication, safeguards against gaming, and stakeholder validation collectively ensure that known threats to validity are proactively identified and addressed.

When interpreting and weighing differing opinions between clinician expert workgroups, review panels, and other interested parties, TEP members highlighted the importance of inclusivity, transparency, and structured evaluation. Many members highlighted giving greater weight to clinicians with extensive real-world experience and specialty societies that represent broader professional expertise, while also valuing patient input for its focus on outcomes and safety. A structured approach prioritizing empirical evidence, applying established evaluation criteria, and adjusting for representativeness and bias was recommended to adjudicate differences fairly. While some favored consensus through voting, others stressed balancing all perspectives equally, recognizing the inherent complexity and uncertainty in cost measure development. One member recommend minimizing the impact of input that conflates cost measure development with quality measure development, as these are inherently different, and cost measures are better suited to identify clear outliers who over-utilize resources. Distinguishing concerns related to cost measurement, program implementation, and measure-specific issues was also noted as critical to fair interpretation. Another member highlighted that there may not always be a definitive answer and stressed the importance of remaining open to diverse perspectives.

2.2.4 Key Takeaways

- TEP members agreed that cost measures are valid when they assess variation in cost performance within the clinician's reasonable influence and measure what they intend to measure.
- TEP members suggested utilizing several different strategies to enhance cost measure validity, including thorough risk adjustment, cost and quality alignment, and comprehensive testing.
- TEP members recommended additional testing to demonstrate that the measure appropriately accounts for patient risk and other factors that are outside of the clinician's control.
- TEP members emphasized the importance of transparent and consistent communication throughout the measure development process, including clear documentation, regular updates to expert panels, public disclosure of methods, and involving diverse stakeholders to build trust and credibility in the validity of the cost measures. Increased communication with the clinician community about cost measures and development efforts could improve clinician understanding.

2.3 Balancing Measure Evaluation Criteria

This session focused on the relationship between validity and reliability, and how to balance trade-offs across these criteria. Section 2.3.1 recaps Acumen's presentation, Section 2.3.2 recaps the TEP members' discussion, Section 2.3.3 summarizes the TEP members' post-meeting survey responses, and Section 2.3.4 contains key takeaways.

2.3.1 Summary of Presentation

Acumen discussed the reliability and validity criteria, which are separate metrics evaluating different aspects of the measure. Validity focuses on accuracy (i.e., whether a measure captures what it intends to capture). Reliability focuses on consistency (i.e., whether a measure produces consistent results under similar conditions). A measure that is valid but not reliable might assess the intended concepts but in an inconsistent way, which would produce results that contain too much noise and have low reproducibility. Meanwhile, a measure that is reliable but not valid might consistently measure unintended concepts, which would produce inaccurate and misleading results. As such, both validity and reliability are important for a measure to yield accurate and consistent data. Reliability standards use a numeric threshold, which allows for a more straightforward answer as to whether or not a measure meets reliability criterion. However, reliability thresholds differ across various guidelines and there is unclear evidence for a selecting a specific threshold. Alternatively, validity is evaluated through qualitative or other subjective criteria. Additionally, there are scenarios in which other evaluation criteria, such as validity, may help justify use of a measure with low reliability, though guidelines are needed to determine how to apply this in practice.

Acumen presented challenges in evaluating reliability and validity together. Although validity doesn't directly reduce reliability, efforts to improve validity can sometimes result in less consistency and impact reliability. Acumen illustrated this tension by highlighting two features that can help measures to focus on costs within the attributed clinician's influence yet reduce variation across providers:

(1) *Use of standardized payments instead of actual payments:* All MIPS cost measures are calculated with payment standardized cost, rather than Medicare payments. Medicare payment amounts reflect factors unrelated to care decisions or delivery, such as geographic price differences (e.g., regional labor costs and practice expenses) and payment adjustments resulting from Medicare programs (e.g., graduate medical education). Payment standardization removes these differences while preserving cost differences that reflect healthcare delivery, such as site of service. Therefore, using payment standardized data allows measures to compare costs on a like-for-like basis; that is, it is a more valid approach to assessing costs than using actual payments. However, using standardized payment instead of actual Medicare payment reduces reliability because removing cost differences generally results in lower between-provider variance and reliability.

(2) *Inclusion of clinically related adverse event costs:* Cost measures assign costs of clinically related services that are within reasonable influence of the attributed provider, which includes adverse events (e.g., hospital readmissions), to provide an assessment of costs of care for an episode or patient. Adverse events often reflect important sources of cost variation that can be influenced by attributed providers through clinical decision making and guideline-concordant care. Including adverse events helps ensure the measure reflects actual resource use and can meaningfully distinguish provider performance. Additionally, including high-cost adverse events aligns with the goals of value-based programs by targeting potentially avoidable and expensive service utilization and incentivizing the reduction of unnecessary spending. Therefore, the inclusion of high-cost adverse events can improve the validity of a cost measure by making it a more accurate, comprehensive, and meaningful reflection of provider performance and resource use. However, as the measure becomes more comprehensive and sensitive to specific adverse events, the variability in provider performance may increase, potentially leading to lower consistency and reliability.

Acumen explained that throughout the measure development process, we aim to develop measures that are valid, reliable, and meet all other measure evaluation criteria. In some scenarios, when it is not possible to strengthen both validity and reliability, Acumen may prioritize validity by (i) gathering other input from experts and persons with lived experience and conducting testing to improve clinical validity, even if it may reduce reliability, (ii) exploring approaches to maximize reliability without compromising measure validity, and (iii) ensuring the

measure meets the program's minimum reliability requirement. Acumen concluded the presentation by noting that it remains challenging to present a holistic view of different measure evaluation criteria (e.g., balancing reliability and validity), in a consistent and transparent way.

Acumen presented the following questions for discussion:

- What do you think about the relative importance of meeting a specific reliability threshold within the current evaluation system, compared to other evaluation criteria?
- What strategies could we use to help evaluation panels and stakeholders develop a shared understanding of the balance between validity and reliability in cost measures?
- How should we approach and communicate situations where a gain in validity might come at the expense of reliability?
- For future TEP meetings, what information would be helpful in reviewing and providing input on measure evaluation criteria?

2.3.2 TEP Member Discussion

TEP members emphasized the importance of ensuring that cost measures are both reliable and valid. Members agreed that there are many factors that influence these evaluation metrics, and while they are related, improved performance on one metric does not necessarily result in improved performance on the other. Generally, members cautioned against prioritizing certain metrics. After reviewing Acumen's examples of reliability and validity trade-offs, members provided feedback on future reliability analyses and underscored the importance of educating clinicians about reliability and validity considerations in cost measurement.

TEP members suggested conducting additional analyses to better understand the relationship between validity and reliability within the context of cost measures. TEP members discussed that impacts to reliability could be due to a variety of factors, including treatment patterns or specific measure specifications (e.g., risk adjustment for disease severity).

Additionally, when discussing the incorporation of standardized payments, TEP members agreed with the need to account for variation in Medicare payments, even if services fall under the same fee schedule, since geographic, regional, and site of service care differences are often outside of clinicians' control. One member noted that there may be some instances cost measures should still reflect differences in care patterns and performance standards that arise from geographic variation (e.g., supplier-induced demand). Members suggested that Acumen look further into identifying what is the influence of treatment patterns compared to other influences that may be creating noise in the measure.

In reviewing the inclusion of readmission costs, members recognized that some adverse events may also be influenced by factors beyond the attributed clinician's reasonable control. Nevertheless, most agreed that these events are important indicators of both cost and quality of

care and should therefore remain included within cost measures. Members recommended additional analyses to evaluate measures where adverse events have a greater impact on reliability in order to better understand the extent of this impact.

Finally, TEP members discussed the importance of increasing education and outreach about cost measure validity and reliability among clinicians. They emphasized the need to strengthen communication about the validity and reliability of cost measures already in use within the MIPS program to help clinicians build a more consistent understanding. They also recommended communicating to clinicians that cost is central to healthcare and can serve as an indicator of quality and that cost represents one dimension of high-value care. One member suggested that when discussing tradeoffs in cost measure validity and reliability, it may be helpful to frame this in terms familiar to clinicians (e.g., the tradeoffs patients face when weighing the benefits and risks of medical care) to encourage greater attention to validity considerations and help narrow the gap between how validity and reliability are weighed in measure evaluation.

2.3.3 TEP Member Survey Responses

In the post-meeting survey, most TEP members agreed that there are scenarios where a measure's lower reliability (below 0.6 but meeting the MIPS minimum of 0.4) could be outweighed by other evaluation criteria, such as validity or importance. Specifically, TEP members identified several scenarios where such limitations could be outweighed by other evaluation criteria, such as high validity, clinical importance, or actionability. Measures addressing high-impact cost areas, rare conditions, or specialties with limited alternatives may be acceptable despite lower reliability, especially when aligned with expert consensus or patient input. Some members emphasized that if a measure is highly valid, narrowly tailored, and not intended for broader use, lower reliability could be acceptable. However, members also underscored that no single criterion should always outweigh another, and that each decision should be made on a case-by-case basis, considering potential harms and benefits. Many members emphasized that reliability below 0.6 should be treated as a concern and suggested that these measures must demonstrate strong correlations with clinical decisions, quality outcomes, or opportunities for quality improvement to justify their inclusion.

When trade-offs arise among measure criteria, such as improving validity at the expense of reliability, TEP members generally agreed that a balanced, context-specific approach is essential. Many members emphasized the importance of prioritizing validity and clinical importance, particularly when measures target outcomes that significantly impact patient care. However, several members cautioned that measures lacking reliability may reduce trust among clinicians, lead to unintended consequences like altered patient selection, or weaken the measure's utility in accountability programs. TEP members advised that measure developers

should clearly document trade-offs, ground decisions in established evaluation frameworks (e.g., CMS, NICE, or National Quality Forum [NQF]), and incorporate iterative testing and stakeholder input.

TEP members emphasized transparency and clear communication as critical when measure developers need to explain trade-offs involving lower reliability balanced against validity or importance. Members noted that measure developers should provide detailed explanations highlighting that lower reliability may result from challenges in measuring complex clinical concepts but that the measure remains the best available indicator. Supporting data demonstrating the absence of adverse effects, such as impacts on patient access or outcomes, can reinforce this communication. Framing the measure within clinical care relevance and maintaining a clear narrative throughout the development process were also mentioned as effective strategies to enhance stakeholder understanding. TEP members recommended using straightforward language, providing explicit analyses of trade-offs, and openly acknowledging limitations while being receptive to future improvements. Engaging stakeholders through group discussions and transparent documentation were also suggested as ways to foster trust and informed decision-making.

To improve clinicians' and stakeholders' understanding of cost measures, TEP members emphasized the importance of transparency and clear communication about how these measures are developed, including detailed specifications and attribution logic. Members stated that clinicians should understand that cost measures reflect observed-to-expected ratios rather than a sum of observed costs, with adjustments made to account for factors outside of their control. Sharing the rationale behind cost measure development, emphasizing their linkage to quality, patient outcomes, and efforts to reduce unnecessary spending, is critical for fostering support. Providing real-world examples, specialty-specific best practices, and plain language explanations about what is included and excluded from cost calculations can further enhance comprehension. Many members highlighted the need for more timely and actionable feedback to clinicians during the performance year, as current annual reports delay opportunities for improvement. Overall, transparency about measure selection, testing, and implementation processes, combined with accessible and ongoing communication, was noted as essential to building trust and encouraging meaningful engagement with cost measures.

2.3.4 Key Takeaways

- TEP members emphasized the need for cost measures to be both reliable and valid, with further analysis required to better understand the relationship between reliability and validity within the context of cost measures.

- TEP members recommended further identifying what is the result of practice patterns and what is the result of outside influences, or noise. Members noted factors like treatment patterns, risk adjustment, and geographic variation can affect reliability.
- TEP member suggested increasing education on cost measure validity and reliability to improve clinician understanding and engagement.

3 NEXT STEPS

Acumen and CMS will carefully consider all guidance received from TEP members during the meeting and in the post-meeting survey responses to inform future measure development, implementation, and maintenance activities. The PCMP TEP will plan to meet again in 2026 to continue discussions on topics related to the prioritization and testing of episode-based cost measures, PRCs, and population-based cost measures.

APPENDIX A: PCMP TEP COMPOSITION LIST

The table below includes the full list of TEP members, their professional roles, and their affiliated professional organizations.

Table A1. PCMP TEP Composition List

Name, Credentials	Professional Role	Organizational Affiliation, City, State
Adolph Yates, Jr., MD	Academic Orthopedic Surgeon	American Association of Hip and Knee Surgeons, Pittsburgh, PA
Amy Aronsky, DO, MBA, FCCP, FAASM	Medical Director	United Healthcare, Princeton Junction, NJ
Barbara Kivowitz, MSW	Patient and Family Advisor	Sutter Health
Barbara Spivak, MD	Practicing Internist, President	Massachusetts Medical Society, Newton, MA
Chloe Slocum, MD, MPH	Associate Chair for Quality	American Academy of Physical Medicine and Rehabilitation, Charlestown, MA
David Kroll, MD	Chair, Committee on Quality and Performance Measurement	American Psychiatric Association, Boston, MA
David Seidenwurm, MD, FACR	Medical Director	American College of Radiology, Sacramento, CA
Denise Morse, MBA	System Executive Director, Quality Analytics	City of Hope National Medical Center, Duarte, CA
Dheeraj Mahajan, MD, MBA, MPH, FACP, CMD, CIC, CHCQM	President and Chief Executive Officer	Chicago Internal Medicine Practice and Research, Oak Park, IL
Gregory Wozniak, PhD	Vice President, Health Outcome Analytics	American Medical Association, Chicago, IL
Jay Nathan, MD	Neurosurgery Quality Council, Chair-Elect	American Association of Neurological Surgeons, Livonia, MI
Jayme Lieberman, MD, MBA, FACS	Vice Chair	Institute for Surgical Excellence, Allentown, PA
Johnnie Sue Wijewardane, PhD, FNP-BC, FAANP	Vice President of Professional Practice	American Association of Nurse Practitioners, Brandon, MS
Joy Gelbman, MD	Assistant Director of Population Health	Weill Cornell Medicine, New York, NY
Karie Nicholas, M.A., G.Dip.	Evaluation and Measurement Manager	Foundation for Health Care Quality, Seattle, WA
Kate Lichtenberg, DO, MPH, FAAFP, FACPM	Medical Director	Anthem Blue Cross and Blue Shield, Leawood, KS
Kevin Klauer, DO, EJD	President and Chief Executive Officer	Shepherd's Hope, LLC, Winter Park, FL
Robert Kropp, MD, MBA, CPHI	Regional Medical Director	American Academy of Neurology, Washington, DC
Rosie Bartel, MA in Educational Leadership	Patient Partner	N/A
Sabrena McCarley, MBA-SL, OTR/L, CLIPP, RAC-CT, QCP, FAOTA, RAC-CTA	Director of Clinical Reimbursement	American Occupational Therapy Association, Napa, CA
Sarah Eakin, MD	President	Pathology Associates of Erie, Erie, PA
Stephen Epstein, MD, MPP	Chair, Board of Governors	American College of Emergency Physicians, Needham, MA
Ugochukwu (Ugo) Uwaoma, MD, MBA, MPH, FACP	Chief Executive Officer	American College of Physicians, South Windsor, CT