Comments on the Pre-Rulemaking Measure Review (PRMR) and Measure Set Removal (MSR) Guidebook

The Acumen Physician Cost Measure and Patient Relationship Codes (PCMP) team thanks the Partnership for Quality Measurement (PQM) for the opportunity to provide feedback on the draft Pre-Rulemaking Measure Review (PRMR) and Measure Set Removal (MSR) Guidebook (June 2026). In Section 1 of our comment, we include a brief discussion of the latest improvements made to the PRMR process covered in this draft Guidebook and potential clarifications that PQM can consider adding. In Section 2, we have reshared our previous comments on several remaining areas for further improvements that we believe PQM should also consider addressing prior to finalization.

1. Improvements and Requested Clarifications

We appreciate updates made to the most recent version of the PRMR/MSR Guidebook, specifically:

- More formal incorporation of the Advisory Group in the PRMR Recommendations Meeting structure; and
- The addition of PRMR Measure Preview meetings to help committee members prepare for PRMR Recommendations meetings, where CMS and measure developers can attend.

In the Guidebook, Section 2.3.4 outlines that Recommendation Group Co-Chairs, along with Battelle staff, will facilitate discussion and ensure that the discussion encompasses the Advisory Group feedback and public comments. However, we have concerns with the ability to have meaningful discussion amongst up to 55-65 committee members during the PRMR Recommendations meeting, who may have varied understanding of the measures and their testing. Typically, a few vocal Recommendation Group members push discussion forward. We request that Battelle consider the following enhancements:

- Provide further clarification on whether the entire Advisory Group's participation in the PRMR Recommendations meeting is mandatory and how the Advisory Group's input will be captured as non-voting discussants during the meetings;
- Formalize how the Committee members' input will be shared prior to and during the Recommendation Group meeting to ensure equal representation across participants and informed discussion; and
- Improve the process for recording and reporting Committee members' votes, rationales, and conditions, which we have elaborated on in Section 2.3.

Our team is very supportive of the PRMR Measure Preview meetings and we believe they will be a valuable opportunity to educate members about measures under consideration and address any questions before recommendation discussions begin. We request that Battelle consider a few additions to this process:

- Request questions from Committee members in advance of the meeting, which encourages Committee members to review measure materials in advance and allows PQM, CMS, or the measure developers to prepare helpful responses and optimize the available time;

- Allow CMS and/or measure developers to provide introductions to each measure undergoing review; and
- Allow CMS and measure developers to provide written responses to the Committee members' questions following the meeting, if we are not able to address all questions within the time we have available or if there is additional information that would be helpful for members.

2. Areas for Further Improvement

During previous public comment periods on the PRMR and MSR Guidebook, we have suggested several improvements to the preliminary assessment process, evaluation criteria, and committee composition and processes. We appreciate where PQM has already incorporated our feedback. We'd like to reshare certain feedback for PQM's consideration, where recommendations may remain applicable.

2.1. Preliminary Assessments

Our comments on the Preliminary Assessments (PAs) focus on the evaluations conducted by PQM staff and committee members (Section 2.1.1) and developer/steward review (Section 2.1.2).

2.1.1. PQM Staff and Committee Member Evaluations

The PRMR/MSR Guidebook provides a high-level explanation of the PAs completed by PQM staff, but does not provide sufficient details for measure developers and committee members to understand the evaluation criteria. The PRMR/MSR Guidebook states that PQM staff evaluate "whether a measure meets criteria related to importance, reliability, validity, feasibility, and usability" (p. 17) and share their evaluation via a PA. It further states that this allows PRMR discussions to focus on whether the measure is appropriate for use. However, the PRMR/MSR Guidebook fails to address several important aspects of this process:

- There is no information about the specific criteria PQM staff use to evaluate "importance, reliability, validity, feasibility, and usability." While one can assume that the evaluation is guided by the Endorsement and Maintenance (E&M) criteria, this is not stated and should be clarified, including which version of the E&M Guidebook is being applied. Additionally, the PRMR/MSR Guidebook should clearly state how PQM staff address issues such as a lack of a consensus process and ambiguity in E&M criteria (e.g., 'not met but addressable' conditions, inconsistent reliability thresholds, lack of clear validity and risk adjustment evaluation criteria).
- There is no information about how Advisory and Recommendation Committee members are supposed to use the information included in the PAs to assess measures' appropriateness. Moreover, there is no information about how committee members can distinguish between concepts with different definitions but similar/identical names in their assessments (see Section 2.2.2 for further comments).

2.1.2. Developer/Steward Review

The PRMR/MSR Guidebook notes that PQM staff develop a PA to assess the measure's importance, reliability, validity, feasibility, and usability, and that measure developers have the opportunity to review them for factual accuracy (pp. 17-18). We ask that PQM continues to allow sufficient time for CMS and measure developers to review the PA. In previous cycles, errors/inaccuracies were not able to be corrected and reposted on the PQM website (and were instead acknowledged verbally at the PRMR/MSR meetings during discussion of the measure). We recommend PQM commit to updating these documents prior to distribution to committee members to ensure information and accompanying analyses are accurately disseminated to committee members.

2.2. Evaluation Criteria

This section outlines our feedback on the relationship between the E&M and PRMR/MSR evaluation criteria (Section 2.2.1), as well as specific feedback on meaningfulness (Section 2.2.2), appropriateness of scale (Section 2.2.3), and time-to-value realization (Section 2.2.4).

Additionally, we would like to emphasize that not all measures are designed to measure quality processes/outcomes or to directly promote quality improvement. For example, cost measures aim to assess cost efficiency, an important focus in today's healthcare landscape and a key component of assessing healthcare value. We recommend that PQM define and explain each evaluation criteria and sub-criteria to refer more broadly to measuring and improving performance, and/or provide guidance specific to each measure type.

2.2.1. Relationship Between E&M and PRMR/MSR

We have previously commented on the potentially overlapping scope of the Endorsement & Maintenance (E&M) and PRMR/MSR process, and requested that PQM provide further clarity, which we did not see incorporated into this version of the PRMR/MSR Guidebook. For example, we noted that during a PRMR Education Meeting, PQM staff stated that the purpose of the PRMR Advisory and Recommendation groups was to review the MUC measures for appropriateness in a CMS program and not to reopen any E&M discussions. We noticed that PRMR committee members have expressed uncertainty around how they should consider the measure testing information, particularly for new measures or measures that have not been submitted for CBE endorsement, where there is no previous discussion or consensus on measure testing to consult. We again ask that PQM provide clarification in the PRMR/MSR Guidebook to ensure that the focus of the discussion is on the appropriateness for use in a program and to clarify how much focus to give on the nuances of measure testing results.

We also note that, unlike the E&M guidebook, the PRMR/MSR Guidebook does not include guidance for committee members to determine whether a submission meets or does not meet a particular evaluation criterion (Appendix B). This presents two major problems: (i) measure developers currently do not have enough information on how to best prepare submissions or be able to develop measures to meet the evaluation criteria, and (ii) this creates the possibility for reviewers to inconsistently apply the evaluation criteria and undermine the transparency in the decision-making process. Currently, readers are limited in the extent to which they can meaningfully provide comments on the evaluation criteria without details about what is considered met or not met. We request that PQM hold an additional comment period once further information is released.

2.2.2. Meaningfulness

The meaningfulness criterion combines several different evaluation criteria into one category, and does not provide guidance to committee members on whether to assign equal weights to the various sub-criteria. Further, we request that PQM update the importance and validity/reliability sub-criteria based on feedback described below.

Importance

Certain measures hold value to the programs as they satisfy congressionally-mandated directives. We would recommend Battelle acknowledge this as one criterion through which the Importance criterion can be fulfilled.

Validity & Reliability

The PRMR/MSR Guidebook creates conflicting definitions of validity and reliability as part of the Meaningfulness criteria. These definitions do not conform to general, scientific understanding of these concepts, to definitions in other Centers for Medicare & Medicaid Service (CMS) guidance

(e.g., the Measures Management System (MMS)), or to definitions in other PQM documents (e.g., the E&M Guidebook). This creates unnecessary confusion for committee members and developers.

The MMS describes validity as follows: “In measure development, 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. Stated more simply, test validity is an empirical demonstration of the ability of a measure to record or quantify what it purports to measure.” In the PRMR/MSR Guidebook, validity is instead defined as "demonstration through data or logic that there are known and effective ways that the person or entity should use to improve the measure focus" (p. 24). We believe that this does not accurately define validity or capture validity. For example, signs reminding staff and visitors to wash hands comply with regulations and lead to reductions in hospital-acquired infections (HAIs). However, a measure that counts such signs is not a valid measure of HAIs or of compliance with all state and federal regulations.

The MMS describes reliability as follows: "A metric is reliable inasmuch as it constantly provides the same result when applied to the same phenomena." In the PRMR/MSR Guidebook, reliability is instead defined as "most of the variation in the measure performance is attributable to variation in the [known and effective ways that the person or entity should use to improve the measure focus]" (p. 24-25). This does not define reliability and does not provide any actionable evaluation metrics. First, there is no clear definition of "most," unless this is intended to be interpreted as "more than 50%." If so, this should be clarified in the PRMR/MSR Guidebook text. Second, in a valid process measure, 100% of measure performance is attributable to person/entity actions by definition, so the criterion is redundant. If the process measure is not valid (such as the example used above), the proportion of variation due to measured actions describes validity rather than reliability. In outcome measures, on the other hand, actions are typically not observed whereas the outcome is, and the measure of the outcome can be more or less reliable. Rather than redefining validity and reliability, we recommend that PRMR (a) focus on importance and usability, and (b) reference MMS or E&M definitions of these concepts. The concept of mechanisms for improvement can be considered as part of Importance (Context of Use) or Usability.

2.2.3. Appropriateness of Scale

The PRMR/MSR Guidebook does not sufficiently explain the Appropriateness of Scale criterion, nor does it provide sufficient guidance for committee members to assess measures on this criterion. The PRMR/MSR Guidebook suggests that this criterion should be an assessment of a measure portfolio and notes the “measure is evaluated in the context of all the measures currently within the program measure portfolio” (p. 4). However, the example PIE questions to support this criterion focus on measure-level benefits and harms rather than program-level context (p. 24 - 25). We request that PQM provide additional clarification about the level at which this criterion should be evaluated. Further, we ask the PQM update the PRMR/MSR Guidebook to support committee members in determining whether the Appropriateness of Scale criterion is met, such as updating the PIE considerations to map more directly to the criterion and providing real-world examples.

2.2.4. Time-to-Value Realization

The Time-to-Value Realization criterion incudes an assessment of whether there is “a plan for near- and long-term positive impacts” as well as consideration of benefits/harms (p. 23). We note that, as currently written, this criterion overlaps with the Appropriateness of Scale criterion. We ask that PQM

consider this overlap as part of our requested updated to the Appropriateness of Scale criterion and provide further distinction between these two criteria. While we recognize that some overlap is inherent in discussion of criteria, having concrete criteria definitions and guidance for committee members is important in ensuring measures are consistently and fairly evaluated.

2.3. Committee Composition and Processes

This section outlines our comments on ensuring sufficient expertise on PRMR/MSR committees (Section 2.3.1) and the committee discussion and voting processes (Section 2.3.2).

2.3.1. Incorporation of Additional Perspectives and Subject Matter Experts (SMEs)

We ask that PQM staff ensure that there is sufficient expertise for the consideration of cost measures during PRMR/MSR. Specifically, PQM could update the PRMR/MSR roster categories (p. 8) to include a target number for members with experiences specific to cost measurement or other specific measure types. Alternatively, PQM should make use of the appointment of subject matter experts (p. 13) to fulfill this important need.

2.3.2. Recommendation Group Voting

We appreciate that committee members who submit a "do not recommend" vote must provide reasoning in writing (p. 29). We would encourage PQM to clarify further in the Guidebook how the reasoning will be recording in writing (e.g., a formal field within their voting poll) to ensure a consistent approach and a detailed written account.

We also recommend the following minor updates to the Recommendation Group voting process to ensure that CMS and developers have sufficient information to interpret the Recommendation Group vote outcome and to inform next steps for further measure development:

- Allow committee members to provide justifications for positive votes, beyond considerations, so that CMS and developers understand why members support a measure for use;
- Have PQM staff provide a pre-determined set of considerations/justifications to promote consistency within the process, and have PQM staff vet any additional considerations/justifications provided by committee members to ensure they align with guidance provided;
- Include indications of the relative frequency of considerations/justifications in summary documents to help CMS and developers understand if a perspective was widely shared or limited to a small set of committee members.

Additionally, we continue to have concerns with focusing discussion on areas of disagreement and minority views (p. 28). We agree that it is important to ensure that all viewpoints receive consideration, which is why we recommend that PQM staff and committee co-chairs also make efforts to note areas where there is consensus among members.