American Hospital Association™
Advancing Health in America

Washington, D.C. Office
800 10th Street, N.W.
Two CityCenter, Suite 400
Washington, DC 20001-4956
(202) 638-1100

July 27, 2023

Partnership for Quality Measurement (PQM)
Battelle
505 King Avenue
Columbus, OH 43201

To the Battelle PQM Team:

On behalf of our nearly 5,000 member hospitals, health systems and other health care organizations, and our clinician partners – including more than 270,000 affiliated physicians, 2 million nurses and other caregivers – and the 43,000 health care leaders who belong to our professional membership groups, the American Hospital Association (AHA) appreciates the opportunity to comment on the PQM's Guidebook of Policies and Procedures for Pre-Rulemaking Measure Review (PRMR) and Measure Set Review (MSR).

AHA has participated extensively in the pre-rulemaking measure review process since its inception as part of the Affordable Care Act. The AHA continues to believe that multi-stakeholder review and evaluation of quality measures before they are used in federal programs can help promote the use of measures that matter the most to providers, patients and communities alike in high-stakes federal programs. By taking on responsibility for supporting the pre-rulemaking process, Battelle must work through a complex transition, while standing up a revised process that all stakeholders find transparent, methodologically rigorous, balanced and credible. We recognize this is a critically important – but daunting – task.

Several of the proposals proffered in the Guidebook have the potential to improve the efficiency of pre-rulemaking review. However, we have quite a few questions and concerns about other aspects of Battelle's plans as well as the overall content of the information provided in the Guidebook. We find the document lacks clarity on certain topics, and there appear to be inconsistencies between statements in the Guidebook and those made on Battelle's own public webinar.

Most importantly, we believe Battelle missed an opportunity to show explicitly how its proposed processes, and evaluation criteria align with the seminal federal policies that govern the use of consensus-based entities (CBEs). Specifically, the National Technology Transfer and Advancement Act (NTTAA) of 1995 and the Office of

Management and Budget Circular A-119 (Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities) together establish the conceptual framework and policies on the development and use of voluntary consensus standards in federal policymaking, and the role of CBEs such as Battelle. Battelle's Guidebook does not make any reference to these policies or how its proposed processes meet the requirements that have been in place for decades. As detailed in our comments on the interested party chapter of the guidebook we are concerned that Battelle proposed approach may not align with the intent of the OMB Circular A-119. **The AHA urges that in the final draft of the guidebook, Battelle make clear connections between these regulatory requirements and the PRMR processes. This step would help to bolster the validity and credibility of the revised pre-rulemaking review process.**

We provide additional detailed comments on specific aspects of PRMR processes, committees, and measure criteria below.

PRMR and MSR Process and Evaluation

In Chapter 3 of the Guidebook, Battelle describes an iterative measure review and evaluation process that combines a Battelle-led assessment with a multi-step process to gather input from interested parties. Battelle suggests this process will "increase engagement of all members and structure facilitation by using standard criteria and practices." The AHA is concerned that Chapter 3 lacks sufficient clarity in exactly how the processes will be carried out. Providing this clarity is especially important because stakeholders had become accustomed to the prior CBE vendor's approach, which naturally leads to questions about how Battelle's process compares to the previous one. In the final Guidebook, the AHA recommends that **Battelle create draft or mock calendars and agendas detailing precisely how and when input would be gathered under this new process.** Without further context, it is difficult to consider how the high-level information presented in the Guidebook will translate into real-life measure review.

Timelines. Battelle proposes that, after the Measures Under Consideration (MUC) list is made available publicly on December 1 of each year, there will be a 21-day public comment period during which Battelle with CMS will hold three public listening sessions. After this period, Battelle will compile information gleaned from public comment and the listening sessions as well as from written feedback from PRMR committees. The Recommendation Group will then meet in mid- to late January.

The AHA believes this timeline is reasonable and an improvement upon that used in the past, where opportunity for public comment and pre-meeting review and analysis was limited. Holding meetings in January rather than mid-December will provide more time for in-depth evaluation of the MUC list, which AHA believes is paramount to meaningful discussion. We also appreciate that CMS plans to hold dedicated public listening sessions, one per "setting" (hospital, post-acute/long-term

care, and clinician); we believe that this system will better allow for robust and relevant public comment.

However, we also have questions about the proposed timeline of PRMR and MSR activities, both for 2023 and for future cycles. For example, the Guidebook notes that the 2023-2024 MSR process has been underway since June and will continue through November. But it is not clear to us how the processes can be underway when the Guidebook outlining those processes is open for public comment through July 30. Similarly, Educational Meetings, described in the Guidebook as those for "PRMR and MSR committee educational meetings, measure developers/stewards, CMS program leaders, etc." are set to begin in August—just two weeks after the end of the public comment period—but there is no information in the Guidebook as to what these educational meetings entail. In addition, committee member appointments start in October, raising the question of who will be able to attend educational meetings if they have not yet been appointed to the committees. Considering the complexity of the tasks involved with developing the MUC list and erecting the structures necessary to review it, it is concerning that Battelle has only provided what amounts to a draft of its plans so close to when the PRMR process is set to begin.

Live Meetings and Voting. In mid- to late January, the Recommendation group meets to discuss issues/concerns raised during the public comment period and feedback from the Advisory group. In the Guidebook, Battelle notes that it will share compiled written feedback from the Advisory group at least two weeks prior to the live meeting to help the Recommendation group prioritize their discussions on areas where consensus is lacking regarding the measures.

Due to the volume of information to be gathered between when the MUC list is made publicly available and when the Recommendation group meets, we suggest that Battelle share a sample of how the information will be presented. Evaluating measures is a complex task and involves coalescing information from multiple sources including not only public comment but also measure specifications, testing results, academic studies, Technical Expert Panel reviews, and past CBE meetings. To ensure that Recommendation group participants are well informed and prepared to thoughtfully engage during discussion, Battelle should make this information accessible and easily navigable (for example, through a web browser with easy-to-follow links between measures and their supporting documentation).

Meeting Facilitation. Recommendation group meetings are facilitated by Battelle staff "to ensure discussions remain productive, within scope, and inclusive of all voices." Previously, workgroup meetings were facilitated by co-chairs appointed by the CBE who were also members of that workgroup. The AHA supports the goals of productive and inclusive discussions. However, we ask Battelle to reconsider using its staff as facilitators rather than members of the workgroup. As we understand it, the primary reason that the previous CBE used co-chairs is that it promoted both the perception and practice of independence from CMS. Staff have always played an invaluable role in

supporting the process and clarifying procedures. However, we are concerned that having facilitate lead the meeting may inadvertently create the impression that they are influencing the trajectory and outcome of the discussion.

Ultimately, the purpose of the pre-rulemaking process is to deliver recommendations that reflect the best thinking and diverse perspectives of the stakeholders that are part of it. We think the best way of meeting this purpose is by empowering the workgroup to facilitate its own discussions, with staff continuing to provide the criteria, boundaries and clarifications that help ensure the process is applied consistently and fairly.

However, if Battelle is intent on using staff as facilitators, it should explain further the safeguards it has in place to maintain independence from CMS. This is especially true given that Battelle's PRMR and MSR work is financed entirely by that federal agency. In addition, workgroup members often have robust institutional knowledge of the CBE processes and measure history that they can deploy while facilitating discussion. While the AHA recognizes that the PRMR and MSR processes are an evolution on the previous process, we hope the Battelle team can find ways to draw in important historical knowledge of measures and the pre-rulemaking process.

Batch Discussion. The Guidebook states that "similar measures" would be discussed in a group "to increase efficiency." The AHA understands the argument for discussing measures with similar specifications to avoid revisiting the same topics *ad nauseum*. However, we request additional information about how the voting and discussion processes will take important differences in measures and in the programs to which they are intended to apply into account.

For example, in the 2022-2023 measure review cycle, the CBE discussed the COVID-19 Vaccination among Patients/Residents measure across various post-acute care settings and came to different conclusions about the measure's appropriateness for inclusion based on the quality reporting program for which the measure was under consideration. That is, the Post-acute Care/Long-term Care workgroup did not support the measure for use in the Inpatient Rehabilitation Facility Quality Reporting Program, but conditionally supported the measure for use in the Skilled Nursing Facility Quality Reporting Program. We assume that Battelle would consider this to be an example of "similar measures;" this was the same measure but under consideration for several different programs. However, we wonder whether a batch of similar measures could also include those that are actually specified differently. For example, in 2022 and 2023, the previous CBE reviewed the Screening for Health-related Social Needs and the Screen Positive Rate for Health-related Social Needs measures for multiple facility-based measurement programs. Stated differently, the nuances in measures that may be "similar" can make enormous differences in their real-life use, so we encourage Battelle to consider how these discussions will take those nuances into account beyond just voting on measures individually.

Quorum. Battelle sets standards for minimum meeting attendance and voting to ensure that discussion and votes are robust and reflective of all perspectives represented in the group. To discuss a measure live, Battelle would require that 60% of the Recommendation group members are present at roll call at the beginning of the meeting; Battelle notes that this number is lower than the threshold of 80% for voting "because of inconvenience and burden of having to reschedule meetings." While we acknowledge the challenges of scheduling, this discrepancy is concerning as it suggests that up to 20% of voters could have entirely missed out on discussion and thus might not be appropriately informed.

We are also confused by the definition of the voting quorum as "at least 80% of active committee members (recommendations group and advisory group)" as, according to other chapters in the Guidebook, Advisory group members do not appear to attend the Recommendations group meetings or vote. We request clarification on this point.

We hope that Battelle will be able to provide more detailed information about the above processes well before the various committees meet. The pre-rulemaking review process is vital in delivering recommendations to CMS to guide federal quality reporting programs. A CBE's credibility rests not just on having the "right" criteria, but also on applying those criteria in a consistent manner each and every time it performs its work. That is why we urge that the final version of the guidebook provide level of detail necessary to determine whether this process will be sufficiently robust and replicable.

Interested Party Organization
The CBE convenes interested parties into committees to participate in the PRMR and MSR. There are three PRMR committees, grouped by care setting: hospital, clinician, and post-acute care/long-term care. Each committee includes two groups of reviewers: the Advisory group, which provides input, and the Recommendation group, which participates in live meetings and provides final consensus recommendations to CMS.

Advisory and Recommendation Group Concerns. Battelle defines the Advisory group as those who "possess a system-level perspective," and includes providers (clinicians and facilities), researchers, purchasers, and other interested parties including but not limited to specialty societies, professional associations (like AHA), EHR vendors, patient safety experts, quality improvement specialists, and national policy makers. According to the information in the Guidebook, this group's participation includes providing written feedback during the PRMR process only. The Recommendation group is defined by Battelle as "those who are most likely to be impacted by the implementation of quality measures," and includes patients, caregivers, patient advocacy groups, providers and facilities, health equity and rural health experts, and purchasers. This group's participation includes both providing written feedback and participating in meetings.

The AHA questions the definitions and makeup of these two groups as there is some but not total overlap. **Furthermore, we are concerned by the inconsistency**

of the information communicated in this draft guidebook with a recent Battelle webinar. In the guidebook, Battelle appears to indicate that every committee member has an opportunity to participate in both the advisory and recommendation groups during their three-year rotation. However, not every committee member would fall under the definitions of each group articulated in the guidebook. For example, Battelle does not list professional associations as those eligible for the Recommendation group. The definitions are inconsistent as well, with *advocacy groups* that work on behalf of members treated differently based on their membership. In the guidebook, Battelle appears to suggest that patient advocacy groups would be considered those most likely to be impacted by the implementation of quality measures, while facility and provider advocacy groups would not.

Certainly, the patients that are members of patient advocacy groups are directly and meaningfully impacted by the quality of care they receive, which can be shaped by the measures used in federal programs. Yet, the statutory basis of the pre-rulemaking process is to provide input on measurement programs that directly affect the reputations and reimbursement of the members of health care facility and professional advocacy organizations. Even more confusingly, a PQM webinar suggested that professional associations *are* considered eligible for the Recommendation group. There is clearly a lack of consistency and clarity in the makeup of these groups; we urge Battelle to provide more explicit information about eligibility for these groups. Battelle must also be consistent in its approach to advocacy groups for patients and health care providers and facilities; we urge Battelle to include both among those that would be eligible for participation in the Recommendation workgroup.

It is worth noting that the American National Standards Institute (ANSI) is private sector, not-for-profit organization that administers and coordinates the US national consensus standards system. It is an authoritative body that can provide advice on how to ensure that the proposed process for developing standards in conformance with the NTTAA and OMB Circular A-119. ANSI can be a useful source of information to ensure compliance with both the letter and spirit of the NTTAA. ANSI committees often include staff from relevant trade associations or professional societies to ensure that the deliberations include not just the voice of a single individual, but rather the voice of an individual whose job it is to represent an entire field or segment of the field or profession.

Committees comprise approximately 60 individuals, and the Advisory and Recommendation groups are mutually exclusive groups of 35-45 individuals and 18-20 individuals, respectively; committee participants are randomly appointed to groups on an annual basis. Battelle argues that this “ensures fairness,” but we question whether this approach ensures that the group making final recommendations to CMS on measures appropriate for inclusion in federal programs will meet the consensus standards in OMB Circular A-119 that recommend gleaning input from a diverse and balanced group of stakeholders. For example, a patient advocacy group could be randomly assigned to the Recommendation group for one year, but to the Advisory

group for two while a purchaser is on the Recommendation group for two years. It is also plausible that the groups review the same or similar measures in multiple years, so in this scenario a purchaser would have multiple opportunities to influence the recommendation on a measure while a patient advocacy group would only have one. This concern is exacerbated by the staggered term lengths beginning with the 2023-2024 cycle: members may only be assigned a term length of a single year, and thus may never get the opportunity to serve on the Recommendation group. **The potential limited time to provide anything more than written input—which can also be done via public comment without any formal participation in these committees—seems to be at odds with the purpose of the PRMR.**

Use of Non-Voting Consultants. Of even greater concern, Battelle notes that it may add "individuals with specialized expertise" to the Recommendation group on an as-needed basis; these individuals would serve as non-voting "consultants." This proposal, which is proffered with little detail, gives us serious pause. We question the purpose of a nomination process to populate these groups if Battelle can choose other individuals to shape the discussion of the measures whenever it chooses. It is unclear how Battelle will identify these consultants, and how it intends to manage any potential conflicts of interest the consultants may have. We have serious concerns that the use of consultants without explicit and transparent criteria and process could introduce inappropriate bias into the measure review process.

Elimination of Coordinating Committee, Health Equity and Rural-Specific Workgroups. Battelle notes that, in the interest of streamlining the number of committees to improve efficiency and fairness, it would integrate the existing Coordinating, Health Equity, and Rural Health committees into the setting-specific committees rather than letting them stand on their own. We do not object to the elimination of an overarching Coordinating committee. However, it is important to note that one of the key purposes of the original coordinating committee was to advise the CBE on broader pre-rulemaking process policies, and to ensure the process was being applied consistently across all workgroups. Battelle may wish to consider engaging a small group – perhaps a working subcommittee of the three workgroups – that it could consult in assessing the consistency of the process across workgroups.

The AHA is concerned that eliminating the opportunity to specifically gather input on how measures will affect rural communities and interact with goals to advance health equity will not result in the same level of consideration of these two extremely important perspectives. The Guidebook suggests that the Recommendation group can include health equity and rural health experts. However, given that participants would be appointed randomly to groups, we are concerned that these two perspectives will not be heard as consistently as they should be.

Volume of Committee Participants. While Battelle argues that the streamlined committees will provide for more efficient and fair review of the measures, the total number of formal participants across all committees is larger than in the previous design

(approximately 180 individuals). We urge Battelle to consider how they will manage this volume in a way that gives meaningful opportunities to contribute to each participant.

Measure Criteria & Evaluation

Staff Assessments. Staff and PRMR participants are asked to assess measure properties based on certain criteria in order to decide whether to recommend the measure for consideration to be added to a Medicare quality program, to recommend it with conditions, or to not recommend it. Committee members are presented with staff assessments and then use those to evaluate or rate the measures based on the evidence therein.

As noted previously, the AHA suggests that Battelle provide a sample of a staff assessment to show what level of information and in what format a measure's properties and evaluation will be presented to participants. The Guidebook asserts that "Committee members must specify and explain if they consulted additional evidence during their evaluation," suggesting it may be unnecessary or even inappropriate to do so. However, considering the diversity of stakeholders comprising the committees, we believe that it is preferable to include as much scientifically valid evidence as possible in evaluating measures for use in federal quality programs in order to capture all relevant perspectives. For that reason, it would be counterproductive to discourage committee members from bringing other relevant studies and data to bear to the discussions.

Furthermore, stakeholders often have important perspectives to share that are informed not by a specific study, but by the totality of their knowledge and experience. This is especially true in the pre-rulemaking process because sometimes measure under consideration are still at the conceptual stage of development and may not lend themselves to a specific scientific study. For example, heath care providers could use their broader experience reflect on whether a measure assesses a relevant topic, has results that would be usable for improvement and the feasibility of data collection. Patient advocates could reflect on whether a measure's results are understandable and useful to them in informing their choice of providers. We urge Battelle to ensure its process does not create undue burdens to raising these important considerations.

Ratings. Participants would be asked to review evidence provided to determine whether a measure meets certain criteria for inclusion in Medicare quality programs. If the evidence that the measure meets these criteria is complete and adequate, a participant would assign the measure a rating of "Recommend," suggesting that the participant recommends the measure for inclusion. If the evidence is either incomplete or inadequate but there is a "plausible path forward," the participant would assign a rating of "Recommend with conditions." If the evidence is either incomplete or inadequate and there is no "plausible path forward," the participant would assign a rating of "Do not recommend."

The AHA appreciates and agrees with these three recommendation categories. The previous cycles of pre-rulemaking review have at times employed categories that

were poorly defined and failed to provide definitive feedback to policymakers. The use of these streamlined categories is more likely to result in consensus.

However, we request that Battelle provide additional information about these categories and how to determine whether a measure has sufficient evidence. For example, the Guidebook does not define what a "plausible path forward" would entail, or what types of "conditions" would be appropriate to recommend. In previous cycles, a common condition for recommendation has been CBE endorsement, as this process can provide clarity on the reliability, validity, and feasibility of the measure's technical specifications. However, because Battelle has severed the endorsement process from the PRMR process, it is unclear whether such a condition would be considered during PRMR or MSR.

Following individual participants assigning ratings to measures, Battelle would identify measures that lack consensus among participants in ratings for further discussion. The Guidebook describes Battelle's "consensus index" that assigns measures a letter category corresponding to whether at least 75% of participants gave a measure the same rating—that is, if more than 75% of participants vote to recommend the measure, it would receive a consensus voting status of "A," whereas if more than 75% vote to recommend the measure with conditions, it would receive a consensus voting status of "B." We are unsure what value this additional categorization provides to the evaluation process and request further clarity on this approach.

Criteria. Participants would be asked to assess whether measures meet certain criteria that recommend them for inclusion. For the PRMR process, there are four criteria. In Appendix B of the Guidebook, Supplemental Guidance on Applying PRMR and MSR Criteria, each criterion is further expanded upon with suggested prompts and consideration for committee members.

The text in Appendix B notes that Battelle "will provide additional guidance to committees about how to apply each criterion. Below we describe some prompts and considerations we could share with committee members to aid in their review." The prompts listed could be helpful, but we wonder whether this format will lead to consistent and meaningful analysis by committee participants. The Guidebook lists several questions a participant might consider in assessing a measure, but does not provide information on how the participant would determine the answer to those questions (e.g., how would one know whether the measure supports best care for all individuals equitably?). We urge Battelle to provide more information about exactly what guidance it will provide, and suggest that this guidance include instructions on how to interpret information in specifications documents and how measure data is reported and used so that all participants—including those coming to the table without a strong background in measure development and evaluation—have an equal opportunity to provide feedback on the measures.

The Guidebook states that the criteria are "intentionally open-ended to allow committees the opportunity to provide holistic feedback about measures under consideration." While we see the merit in assessing measures holistically rather than getting lost in the minutiae of statistical performance, the lack of specificity in the criteria and the supplemental guidance is troubling. We fear that this "open-ended" approach could lead to inconsistently applied standards for measure usefulness, leading to recommendations for measures that may address important issues but do so in a way that does not lead to improved patient outcomes when the measures are deployed. We provide specific considerations for the criteria below.

Meaningfulness. Battelle defines this criterion as "importance, feasibility, scientific acceptability, and usability...considering the use across programs and populations." The previous CBE considered each of these characteristics individually, often with specific minimum statistical acceptability standards. We worry that combining all of these characteristics into a single category rather than evaluating them individually will result in less rigorous analysis.

Appropriateness of Scale. The Guidebook notes that participants should consider how "implementation of the measure [is] applied to optimize the measure value across segments of the target population and entities of the program under consideration." We request clarification about the meaning of this goal, and how a committee participant is intended to evaluate a measure against this criterion. Similarly, participants are encouraged to consider whether "the measure may identify an equity gap." Measures do not inherently identify gaps; rather, the data generated by collection and analysis of performance on the measure might assist providers, researchers, and other stakeholders identify areas for improvement. The use of quality measurement to advance equitable health outcomes is an important goal of Medicare quality programs, and we encourage Battelle to give more attention to this consideration than a confusingly worded mention in passing.

Time to Value Realization. Many of the considerations within this criterion appear to overlap with those in the others. For example, "a clear pathway from measurement to performance improvement" seems to be the same characteristic as "progress in this measure demonstrably improve[s] care," a consideration in the Meaningfulness criterion. Other prompts are purely theoretical, meaning they are not based on available evidence and would thus be challenging for participants to evaluate. For example "How might measurement support the generation of better evidence in the future? How might that evidence mature over time to reduce uncertainty about how entities may best improve outcomes?" It is not clear to us what information or evidence a participant would cite to answer those questions. We urge Battelle to rethink this criterion to define it more specifically and distinctly from the others in order to provide useful information about the measure's appropriateness for inclusion in Medicare quality programs.

Finally, we are concerned that the PRMR criteria appear to be missing some important aspects of quality measure evaluation. For one, we believe that participants should

consider—even at a high level—measure testing performance as an indicator of whether the measure is likely to achieve the objective for its use. While we understand there are other bodies within PQM that will review the scientific merits of individual measures, it remains important to consider whether a measure is an accurate representation of performance rather than a variable indirectly associated with performance, influenced by forces outside of the provider's control. Similarly, participants should consider whether there is an opportunity for improvement based on testing results; a measure that will be adopted into a Medicare quality program already "topped out" will not result in substantive improvements in patient outcomes.

These considerations help differentiate measures that merely mention an important topic from those that will hold providers accountable and help improve care. In other words, we all agree that quality measurement is a worthwhile undertaking—we believe it is worth doing right.

Again, we thank you for your consideration of our comments. Please contact me if you have questions or feel free to have a member of your team contact Caitlin Gillooley, AHA director of behavioral health and quality policy, at (202) 626-2267 or cgillooley@aha.org.

Sincerely,

/s/

Nancy Foster
Vice President
Quality & Patient Safety