Impact Assessment of CMS Quality and Efficiency Measures
Contract #: 75FCMC18D0026; Task Order #: 75FCMC19F0001

CMS Measure Development Plan (MDP)/ Quality Measure Index (QMI) Technical Expert Panel (TEP) Meeting Summary

Meeting Date: April 15, 2024

To: Nidhi Singh Shah, MPH, QMI Project Lead
7501 Wisconsin Ave., Mail Stop 8-008
Bethesda, MD 20814-6519
Nidhi.Singh-Shah@cms.hhs.gov

Michelle King, MASS, BS, COR III
Contracting Officer's Representative
Centers for Medicare & Medicaid Services (CMS)
7500 Security Boulevard, Mail Stop S3-02-01
Baltimore, MD 21244-1850
Michelle.King@cms.hhs.gov

Table of Contents

Technical Expert Panel Meeting Summary: MDP and QMI Project Updates

Executive Summary

Meeting Overview

On April 15, 2024, the Health Services Advisory Group, Inc. (HSAG) team convened a meeting of the CMS Measure Development Plan (MDP)/Quality Measure Index (QMI) Technical Expert Panel (TEP). Twelve of 15 TEP members attended at least a portion of the meeting.

Meeting Objectives

- Provide update on MDP project activities since the October 2022 meeting.
- Provide update on QMI project activities since the March 2023 meeting.
- Review refinements to QMI Tool Version 2.1.
- Obtain TEP input on the application of the QMI tool to measures under development (MUDs).

Key Takeaways from TEP Discussion

- General
  - TEP members asked whether the QMI tool methodology is available to measure stewards. The team responded that the tool is currently an internal CMS tool but that data elements and definitions from the QMI are included in the Measures Under Consideration (MUC) Entry/Review Information Tool (MERIT) and guidance.
- Agency High Priority domain
  - *Preferred Measure Type:*
    - The team clarified that the Universal Foundation component of the Preferred Measure Type variable reflects the measure developer's attestation of whether a measure is currently included in the Universal Foundation (but is being submitted for a different program).
- Importance domain
  - *Evidence-based:*
    - A TEP member suggested requiring measures to illustrate they are actionable. TEP members recognized the value of the QMI assessing whether a measure submits a logic model depicting relationships between structures, processes, and the desired outcome in the future. Members expressed concern it might be burdensome for measure developers to demonstrate actionability in a different way, such as a randomized trial.

- Scientific Acceptability domain
  - *Measure Score Reliability:*
    - A TEP member asked whether the QMI tool has a threshold for sample size. The team responded that currently it does not, as the team wants to avoid defining QMI criteria more strictly than criteria defined by the consensus-based entity (CBE), which has not established such a threshold.
  - *Validity (composite):*
    - A TEP member expressed concern that the requirement to test measures in at least two EHR systems is burdensome for developers. The team responded that this requirement has generally been the expectation for eCQMs from the CBE.
- Feasibility domain
  - *Feasibility of Data Elements*
    - A TEP member asked whether artificial intelligence (AI) has been explored as an option to search text and pull information out of unstructured data. The team responded that the variable score is determined simply based on measure developer attestation in MERIT of whether data elements are available electronically in defined fields. The team is monitoring the use of natural language processing (NLP) in quality measures and will consider updates in the future as the field evolves.
  - *Provider Burden (Impact on Workflow)*
    - A TEP member noted that if desired practices were routine, we wouldn't need quality measures for them. Although a measure may create a new workflow, the measure may make meaningful changes in care delivery. The team acknowledged that CMS has to consider this tradeoff when a preferred measure creates more burden for providers. The QMI tool shows measure strengths and weaknesses.
- Measures under development (MUDs)
  - Eleven TEP members approved the proposed scoring approach for MUDs that have completed the initial specification stage of development and have not begun beta testing. There were no suggested modifications to the proposed approach.

Next Steps

This was the last meeting for the 2022–2024 project period; TEP membership will end June 30, 2024.

Introduction

The Centers for Medicare & Medicaid Services (CMS) has contracted with Health Services Advisory Group, Inc. (HSAG, "the team") to develop and maintain the CMS Measure Development Plan (MDP) and the Quality Measure Index (QMI) tool under Contract #75FCMC18D0026; Task Order #75FCMC19F0001. As part of this contract, HSAG convenes a CMS MDP/QMI technical expert panel (TEP) composed of patients and family caregivers, clinicians, representatives of professional societies, consumer advocates, and quality measurement experts to provide input on project tasks and reports. HSAG convened an MDP/QMI TEP meeting on April 15, 2024, to provide project updates and obtain TEP input on proposed methodology.

Meeting Proceedings

Opening Remarks and Objectives

Presenter: Julia Mackeprang, MPH, PMP; Project Lead, QMI

J. Mackeprang welcomed the TEP members. After reviewing ground rules, she reviewed the agenda and meeting objectives. Meeting objectives were as follows:

- Provide an update on MDP project activities since October 2022 TEP meeting
- Provide an update on QMI project activities since March 2023 TEP meeting
- Review refinements to QMI tool version 2.1
- Obtain TEP input on application of the QMI tool to MUDs
- Review the project timeline and next steps.

TEP Roll Call and Conflict of Interest Disclosures

Presenter: Samantha Tierney, MPH; Senior Scientist, Clinical Policy, American College of Physicians; TEP Co-Chair

S. Tierney conducted roll call. Twelve of 15 TEP members attended at least part of the meeting**.**

TEP members disclosed the following conflicts of interest:

- **Z. Butt** receives equity from his employer, Phreesia, which stewards a measure for the Merit-based Incentive Payment System (MIPS). He noted that he is located in Illinois, but Phreesia's home base is now in Delaware.
- **C. Eppes** receives funding from CMS for the Medicaid Maternal Opioid Misuse Model.
- **N. Gupta** is on a medical advisory board for home dialysis at Fresenius Medical Care.
- **G. Miller** noted that Vituity®, his employer, sponsors a qualified clinical data registry (QCDR) and has several measures.
- **A. Somplasky** is employed by Mathematica Policy Research, a CMS contractor that oversees annual updates for MIPS measures.

Invited Members and Guests/Attendance

TEP Members	CMS	HSAG
☒ John Martin *(Co-Chair)*	☐ Nina Heggs	☒ Kyle Campbell
☒ Samantha Tierney *(Co-Chair)*	☒ Michelle King	☒ Eric Clark
☒ Mary Baliker	☐ Helen Dollar-Maples	☒ Hayley Dykhoff
☐ Crystal Barter	☐ Nidhi Singh Shah	☒ Mariana Grass
☒ Heidi Bossley	☒ Marsha Smith	☒ Eric Gilbertson
☒ Zeeshan Butt	☒ Mei Zhang	☒ Nancy Gordon
☒ Catherine (Carey) Eppes		☒ Marie Hall
☒ Nupur Gupta		☐ Susan Hemmingway
☒ Amy Nguyen Howell		☐ Megan Keenan
☒ Shu-Xia Li		☒ Julia Mackeprang
☒ Gregg Miller		☒ Michelle Pleasant
☐ Connie L. Montgomery		☒ Rob Ziemba
☒ Sarah Hudson Scholle		
☒ Anita Somplasky		
☐ Lindsey Wisham		

Update on MDP Project Activities

Presenter: Michelle Pleasant, PhD, MA; Project Lead, MDP

M. Pleasant reviewed progress of the MDP project since the October 2022 TEP meeting, noting that the project team has since produced and publicly posted the 2022 and 2023 MDP Annual Reports and that the 2024 MDP Annual Report is in federal clearance for public posting by May 1, 2024.

M. Pleasant discussed highlights of the 2024 MDP Annual Report, which included total measure development activities, spending, and the number and type of measures developed for fiscal year (FY) 2023 as well as cumulatively for FY 2016–2022.

She continued that the 2024 MDP Annual Report will be the last in a series of eight annual reports due to expiration of funding. CMS will continue to monitor measure development for the Quality Payment Program internally.

TEP members had no questions or comments.

Update on QMI Project Activities

Presenter: Julia Mackeprang, MPH, PMP; Project Lead, QMI

J. Mackeprang reviewed progress of the QMI project since the last TEP meeting in March 2023. This includes extending the QMI tool to health plan–level and structural measures; scoring 39 measures submitted to the CMS Measures Under Consideration (MUC) Entry/Review Information Tool (MERIT) in 2023; analyzing 2023 MERIT data to refine QMI tool version 2.1 for 2024; and proposing refinements to QMI methodology to the TEP workgroup and the consensus-based entity (CBE).

J. Mackeprang recognized the 10 TEP members who served on the QMI workgroup to provide feedback on the proposed QMI methodology refinements.

QMI Tool Version 2.1: Structure and Scoring Approach

Presenters: Julia Mackeprang, MPH, PMP; Project Lead, QMI, and Hayley Dykhoff, MSPH, Lead Analyst, QMI

J. Mackeprang reviewed the structure of QMI tool version 2.1 and the tool's scoring approach. H. Dykhoff then displayed sample QMI score worksheets to show how QMI classification variables are used and measure scores are represented in a QMI Scores Workbook that summarizes each measure's QMI score results. H. Dykhoff also shared a sample QMI Scores PDF Report, which is a two-page summary report produced for each scored measure containing measures information including QMI scoring information.

TEP Discussion

- A TEP member asked in a chat message whether QMI information is available to measure stewards.
    - K. Campbell (HSAG) replied QMI is presently limited to CMS internal use only.
- J. Mackeprang acknowledged another TEP member's question in a chat: Is the expectation that the domains will always be equally weighted in deriving the overall QMI score? She replied that the domains are currently equally weighted; however, the team is continually evaluating the tool for refinements.
    - H. Dykhoff (HSAG) followed up by asking whether the questioner had thoughts about ways the domains might be weighted.
    - The member had no alternative to propose, but expressed curiosity about whether the decision was based on conceptual grounds or was data driven.
    - K. Campbell (HSAG) explained that the rationale was conceptual. The team spent some time considering the weighting, given that various interested parties might choose to assign different weights to certain domains.

Review Refinements to QMI Tool Version 2.1

Presenters: Julia Mackeprang, MPH, PMP; Marie C. Hall, RN; Mariana Grass, DHSc, MPH; and Hayley Dykhoff, MSPH

J. Mackeprang began review of QMI tool version 2.1 updates by reviewing the key objectives for QMI tool refinements.

Agency High Priority Domain: *Measure Focus, Digital Data Source,* and *Preferred Measure Type*

M. Hall introduced the three variables in the Agency High Priority domain, showing their updated operational definitions, how they are scored, and modifications agreed upon by the TEP workgroup and CMS. In discussing the *Measure Focus* variable, she noted two new criteria were added for this year, including the Meaningful Measures 2.0 priority of Person-Centered Care and whether the measure is stratified to address an equity gap based on evidence from testing and/or

the literature. For the *Digital Data Source* variable, she showed the CMS definition of digital quality measures (dQMs) and digital data sources for dQMs. For *Preferred Measure Type*, she noted CMS' request to incorporate the Universal Foundation of Measures as an added variable criterion.

TEP Discussion

- A TEP member complimented the team on the slides and asked about adding the Universal Foundation as a criterion: Is that for whatever measures are in the Universal Foundation today? My understanding is the Universal Foundation is supposed to evolve over time.
    - M. Hall (HSAG) replied: We understand CMS will be updating the Universal Foundation of Measures. The QMI tool will score the measure at the time the developers submit a measure to MERIT, when they attest to whether the measure is a part of the Universal Foundation. If the measure is included at the time of submission to MERIT, they receive a Green score for this variable.
- The member had a follow-up question: Is it correct that it captures whether a measure has already been named in the Universal Foundation, and not that it could be consistent with a domain in the Universal Foundation?
    - M. Hall (HSAG) agreed: Right, it could be in the Universal Foundation but not necessarily in the program, and the measure would get credit for that. The reasoning is that there may be measures in the Universal Foundation that are not necessarily the preferred measure types of outcome, intermediate outcome, or PRO-PM [patient-reported outcome performance measure].
- Recalling the issues with weighting and noting duplication across measure domains, a TEP member questioned whether a measure in the Universal Foundation was getting double credit.
    - M. Hall (HSAG) replied: If the measure were already in the Universal Foundation, we would give it credit in the sense that it would score Green, but it's part of giving the measure a score for meeting the Agency High Priority domain.
    - K. Campbell (HSAG) explained the rationale: If a Universal Foundation measure is listed for an existing CMS program, applying the same measure in another program would promote alignment, which is what CMS is hoping to see.
    - Regarding the expansion of the Universal Foundation, K. Campbell mentioned the recent addition of three measure sets for hospital, post-acute, and maternity care. As the Universal Foundation evolves, these measures get identified if they are in one program but not in another.
- Another member noted that the adult set is already in MIPS, but she understands now that those measures could be proposed for another program.
    - K. Campbell (HSAG) concurred: For example, if [MIPS measures] were going to be used in Part C or D Star Ratings or Medicare Shared Savings or other programs, that's where they would get credit.
    - H. Dykhoff (HSAG) pointed out that the *Preferred Measure Type* variable encompasses measure type as well as Universal Foundation measures.

- M. Hall (HSAG) expanded on that point: If a measure doesn't meet a preferred measure type, we check to see if it is part of the Universal Foundation and give credit to those measures, many of which are process measures.
- The team shared a link to the CMS Universal Foundation of Quality Measures webpage: https://www.cms.gov/medicare/quality/cms-national-quality-strategy/aligning-quality-measures-across-cms-universal-foundation

Importance Domain: Evidence-based, Measure Performance, and Patient Engagement

M. Grass began review of Importance domain variables with *Evidence-Based*. She showed how its operational definition differs for scoring outcome measures versus other measure types and noted there were no meaningful changes to the variable since the prior version of the QMI tool.

She reviewed the *Measure Performance* variable, its updated operational definition and scoring criteria, and how the order of those criteria changed in scoring. For the *Patient Engagement* variable, M. Grass showed the updated operational definition and scoring that assesses whether the majority of patients/caregivers consulted agree that the measure is meaningful and valuable. She noted that the TEP workgroup agreed to the revised operational definition.

M. Grass presented the following question for TEP discussion:

> Are there any additional objective criteria that could be considered to strengthen the Importance domain while imposing limited burden to measure developers?

M. Grass asked for specific feedback: The CBE requires measure submitters to attach a logic model depicting relationships between structures, processes, and the desired outcome. The QMI does not require that the logic model illustrate such relationships. She asked: Is that something that the TEP would like to see? Or is the *Evidence-Based* variable's attestation of a relationship itself is sufficient?

TEP Discussion

- A TEP member proposed to strengthen the QMI criterion by requiring an illustration that a measure is actionable. That amendment would apply to all types of measures.
- The member asked whether the requirement could apply to a process measure, or if actionability is referenced in the guidelines or the evidence base.
  - K. Campbell (HSAG) agreed the team could consider having developers show in the logic model how the process relates to the outcome for all measure types.
- Another TEP member appreciated the idea of the logic model but expressed concern that demonstrating actionability in a different way might be a burden to measure developers.
- A third TEP member agreed. Speaking from experience in research and measure development, she noted that the pathway to reach a measure outcome is complex.
- The first speaker clarified her intent: The logic model is less about showing you improved performance of a measure to this level of confidence with, for example, a randomized trial. It's more about, based on existing evidence, there are ways to improve the outcome. The logic model tells the story based on existing evidence, that there is something that you can do and why it would be relevant to this accountability entity.

What I was looking for in the logic model is that you're able to say there are interventions that work toward the intent of the measure, not the specific performance on the measure.

- The explanation alleviated some concerns of a previous speaker, who commented: I do think most of that is theoretical—if you improve prescription rates, then you'll have better outcomes, better performance, [and] reduce comorbidities. But it wasn't ever based on evidence; it was the thought process that our committees would have. I certainly agree with you: If a measure is not actionable, then why is it even a measure?

Scientific Acceptability Domain: *Measure Score Reliability and Validity (Composite)*

H. Dykhoff introduced the Scientific Acceptability domain variables, starting with *Measure Score Reliability*. There were no meaningful changes for *Measure Score Reliability*.

TEP Question

- A TEP member asked the presenter to explain what "invalid sample size" means, with regards to scoring Grey for the *Measure Score Reliability* variable.
    - H. Dykhoff (HSAG) responded that the QMI tool treats a value as invalid if it does not make sense to represent a sample size—such as a negative number, one with a decimal point, or the figure 9999, intended to convey "not assessed" or "unknown."
- The member asked whether there is ever a determination that the sample size is inadequate to support reliability.
    - H. Dykhoff replied that the team has considered adding more robust parameters on sample size to the QMI tool. However, the team wants to avoid defining QMI criteria more strictly than criteria defined by the CBE, which has not established strict boundaries or thresholds on what constitutes a meaningful sample size.

H. Dykhoff resumed the presentation to discuss the *Validity (Composite)* variable, noting *Validity* is a composite of four components: Empiric Measure Score Validity, Data Element Validity, Face Validity, and Risk Adjustment. The Data Element Validity component of the *Validity* composite added the criterion of requiring eCQMs to be tested using EHRs from at least two EHR vendor systems.

TEP Question

- A TEP member noted that only face validity is required for measures on the MUC List to go into MIPS for the first year. She wondered whether such measures would score Red for the other three component variables (Empiric Measure Score Validity, Data Element Validity, and Risk Adjustment), or Green because they had what was required.
- The member also noted that MIPS requires risk adjustment only for cost measures and those based on administrative claims. So, she asked, is the plan to keep Risk Adjustment in the tool, even though we know it's not required?
    - H. Dykhoff (HSAG) displayed the QMI scoring algorithm for the *Validity* (Composite) variable, explaining that risk adjustment is assessed only for

outcome, intermediate outcome, and PRO-PMs. Further, a measure does not need to be risk-adjusted or stratified to obtain a Green score on the Risk Adjustment component: If an adequate rationale for not risk-adjusting is provided, a measure can also get a Green.
    - K. Campbell (HSAG) added that for QCDRs, the understanding is that face validity is acceptable regardless of the type. But for traditional MIPS, for example, if it's an eCQM, our understanding is the expectation still would be around that comparison between the electronic extract and manual review of the record, perhaps for data element validity—which is the way the QMI essentially enforces that in the algorithm. He encouraged the TEP member to share her thoughts about that requirement for traditional MIPS.
    - The member agreed that eCQMs are being held to that higher standard in terms of what they need to show for empiric measure score validity. However, the QCDRs are still just showing face validity for their first year.
    - K. Campbell responded that this was consistent with the QMI algorithm. He clarified for the rest of the TEP that QCDRs are in a different pathway than MERIT and MUC, so the QMI tool is not used to score these measures. At this phase, we're applying the QMI tool to measures submitted to MERIT through the pre-rulemaking process.
    - The TEP member questioned whether QCDR measures will be held to the same standard in the future, as MIPS Value Pathways move forward and QCDR measures are included.
    - K. Campbell responded that the team would defer to CMS in terms of potential application of the QMI to QCDR measures in the future.

H. Dykhoff (HSAG) resumed the presentation. For the Data Element Validity component, a criterion was added to require eCQMs to be tested using EHR data from at least two EHR vendor systems; the TEP workgroup had previously agreed to the change.

H. Dykhoff presented the following question for TEP discussion:

> Are there any additional objective criteria that could be considered to strengthen the Scientific Acceptability domain while imposing limited burden to measure developers?

The TEP co-chair moderating the discussion invited TEP members to comment on the requirement for measure testing in two EHR systems.

TEP Discussion

- A TEP member contended that the testing requirement is overly burdensome for developers, particularly since the Measure Authoring Tool (MAT) will be no longer available at the end of June.
    - M. Hall (HSAG) added that everything is transferring to MADiE [Measure Authoring Development Integrated Environment].
    - K. Campbell (HSAG) observed that testing with a minimum of two EHR vendors has generally been the expectation for eCQMs. He suggested that the TEP

member may be alluding to the burden for developers to transition eCQMs to the Fast Healthcare Interoperability Resources [FHIR] standard. He explained that the FHIR transition isn't something the QMI team has operationalized as a requirement yet, but it may be something to consider for a future version of the QMI tool.

- Another TEP member acknowledged the added burden of testing with two EHR vendors but agreed that it is important to have at least a couple of vendors. He added that sometimes small-scale vendors that are selected may not represent a broader use—although nothing is wrong with their systems.

Feasibility Domain: *Feasibility of Data Elements* and *Provider Burden (Impact on Workflow)*

M. Hall reviewed the Feasibility domain, showing the operational definitions and scoring approaches for the two variables, *Feasibility of Data Elements* and *Provider Burden (Impact on Workflow)*. She noted that no meaningful changes were made to either variable.

M. Hall presented this question to the TEP for discussion:

> Are there any additional objective criteria that could be considered to strengthen the Feasibility domain while imposing limited burden to measure developers?

She noted that the CBE has no criterion labeled as "burden" but considers data element feasibility as a path to support electronic data capture. Alternatively, to meet the criterion, the measure data can be routinely generated and used during the course of care—which mirrors what the QMI captures under the *Provider Burden* variable.

TEP Discussion

- A TEP member asked whether there had been any thought about the potential for artificial intelligence (AI) to search the text and pull information out of unstructured data.
  - H. Dykhoff (HSAG) replied that some interesting discussions have occurred in a monthly group she attends about how developers can use natural language processing (NLP) in their algorithms to search for specific text, terms, or phrases in the clinical notes. Then they enter an ICD-10 code that matches the phrase they found in a defined field. This work is evolving and will need to be standardized in the future.
- The member followed up: That's exactly what I had in mind. Does that count? As long as it gets into a defined field, it doesn't matter if a human did it or if a computer did it; it's OK? It just has to be in a defined field—is that the rule?
  - H. Dykhoff replied: This is a measure submitter attestation question. There are no hard and fast rules.
  - M. Hall added: The QMI automates the score by taking exactly what is submitted to MERIT. She agreed with the idea of reviewing the guidance to submitters during the next MERIT update.

- The member observed: You're going to have different people interpret it differently, so if you want comparable data, more instructions could be helpful. It's probably worth putting on a list of questions to ask to inform future versions.
- The member commented further about a measure's potential impact on workflow. If [a desired practice] were routine, we wouldn't need a quality measure for it. You know a measure that is based on patient data is going to create a new workflow, but they get weighted extra somewhere else. I see that CMS is trying to balance it out, and I'm trying to understand how much this may hold us back from developing measures that are designed to make meaningful changes in how care is delivered. We're trying to follow the evidence rather than what people [already] do today.
    - M. Hall agreed that CMS has to consider a tradeoff when a preferred measure type creates more burden for clinicians and providers.
- In closing, the member compared submitting measure information to MERIT with the process for seeking endorsement: Measure developers can see in the CBE application how that [information] gets represented. But in the case of the QMI, developers don't ever get the feedback. It's not so much so that people can game the questionnaire, but more that they understand how CMS is using this tool to evaluate measures. I realize that the tool is for CMS now. I'm curious about the logic of how CMS uses it and how CMS might share it with measure developers for transparency and in their decision-making.
    - K. Campbell assured the member that the team will pass those comments on to CMS. He noted that the QMI is not the only tool CMS uses in measure selection but acknowledged potential benefit from having more information publicly available. He added that all of the guidance and information related to the QMI variables is baked into MERIT, and that information is public. CMS has a developer orientation on the use of MERIT, and the QMI team shares information about expectations for each of the objective variables that are part of the QMI.
    - J. Mackeprang added that the team tries not to define QMI criteria more strictly than the CBE criteria or the Measures Management System guidance.

Proposed Application of QMI Tool to Measures Under Development (MUDs)

Presenter: Mariana Grass, DHSc, MPH; Health Services Researcher, HSAG

M. Grass reviewed the stages of the Measure Lifecycle, noting that the team has proposed an approach to score MUDs that have completed the initial specification stage of development and have not begun beta testing. She presented the benefits of scoring MUDs, the proposed MUD scoring methodology, and an example of a measure full QMI score versus a score calculated using the MUD scoring methodology. The methodology for MUDs omits the variables *Measure Performance*, *Measure Score Reliability*, and *Validity* because MUDs lack the data required to score these variables.

M. Grass presented the following question for TEP discussion:

> Do you agree with the proposed scoring approach for MUDS?

She noted that the presented scoring approach avoids penalizing measures that would otherwise not have measure performance or scientific acceptability testing data at this stage of the measure lifecycle. It simplifies implementation of the QMI tool by aligning the overall QMI scoring approach across lifecycle stages, and it reduces data collection burden for variables that are not yet applicable to MUDs. In the future, the team could consider adapting the *Measure Performance* variable to assess an anticipated performance gap based on the literature.

TEP Discussion

- A TEP member wondered whether this would be applied to measures that are not quite fully developed but are submitted to MERIT.
    - M. Grass confirmed that the presented approach is only for CMS measures that are under development. The QMI scores could be used to inform a CMS decision to move the measure to the more costly stage of testing, or the measure might need additional help before it reaches that stage of testing. Measures submitted to MERIT are all scored using the full QMI scoring methodology because of the expectation that those measures can be used in CMS programs. The MERIT website encourages the submission of fully developed measures.
- Another TEP member asked how the QMI currently scores a measure that has not been fully tested.
    - J. Mackeprang (HSAG) clarified that for a MERIT submission this year, the measure would be scored using the same algorithm as measures that are fully developed. The application we're discussing here applies to measures that are under development to help CMS decide whether they want to continue development. Specifically, whether CMS wants to move into beta testing, which is the costliest stage of measure development.
- The first TEP member questioned why the team is considering adapting the performance variable for MUDs in the future, not now.
    - M. Grass (HSAG) stated that the team intends to discuss with the CBE whether it would be acceptable to base an assessment of the performance gap on literature; this information will help to inform potential refinements to the QMI methodology.
    - K. Campbell (HSAG) added: We're working to get the QMI on an annual update cycle for maintenance, whereby all the applications of the QMI align. As we develop the version applicable to MERIT, we wouldn't make any changes to the tool in advance of the next cycle. So, the tool is aligned across the various stages of the lifecycle in a given year.

TEP Vote on Proposed MUD Scoring Approach

Eleven members approved the MUD scoring approach with no changes, including 6 members who voted during the meeting and 5 members who voted after the meeting.

Timeline and Next Steps

Presenter: Julia Mackeprang, MPH, PMP; Project Lead, QMI

J. Mackeprang showed the project timeline for the QMI from January through June 2024 and next steps for the TEP. She noted that this TEP meeting is the last for the 2022–2024 project period and that TEP membership will end June 30, 2024.

J. Mackeprang concluded the call by thanking the TEP members for their participation and engagement and thanking CMS for their support of the project.