CMS Quality Measure Development Plan (MDP)/
Quality Measure Index (QMI) Project
Contract #: HHSM-500-2013-13007I Task Order #: HHSM-500-T0002

Technical Expert Panel Meeting Summary Meeting Date: June 17, 2019

MACRA Section 102

Submitted: June 25, 2019

To: Noni Bodkin, PhD, Contracting Officer's Representative (COR)
Centers for Medicare & Medicaid Services (CMS)
7500 Security Boulevard, Mail Stop S3-02-01
Baltimore, Maryland 21244-1850
Noni.Bodkin@cms.hhs.gov

Table of Contents

Technical Expert Panel Meeting Summary

I. INTRODUCTION

The Centers for Medicare & Medicaid Services (CMS) contracted with Health Services Advisory Group, Inc. (HSAG) to develop the *CMS Quality Measure Development Plan: Supporting the Transition to the Merit-based Incentive Payment System (MIPS) and Alternative Payment Models (APMs)*[1] under Contract #HHSM-500-2013-130071; Task Order #HHSM-500-T0002. As part of this contract, HSAG ("the team") is also tasked to develop the CMS Quality Measure Index. HSAG convened a multidisciplinary technical expert panel (TEP) of stakeholders (e.g., patients and family caregivers, clinicians and representatives of professional societies, consumer advocates, quality measurement experts, and health information technology specialists) to gather their recommendations on options for weighting the variables included in the Quality Measure Index.

II. BACKGROUND

On June 17, 2019, HSAG convened the fourth meeting of the 2018–2019 Measure Development Plan (MDP) TEP by webinar. The meeting's key purpose was to provide updates on the beta testing results of the Quality Measure Index and solicit TEP input on weighting strategies for the index. Seventeen of 23 TEP members attended, along with HSAG staff. Present from CMS were Noni Bodkin, Contracting Officer's Representative; Nidhi Singh Shah, Project Lead; and Wilfred Agbenyikey, Health Insurance Specialist. The objectives of the meeting were to:

- Provide an update on MDP-related activities since the previous TEP meeting.
- Present the Quality Measure Index workgroup's achievements and recommendations.
- Review developments in Quality Measure Index beta testing.
- Recommend a Quality Measure Index weighting scheme.
- Discuss next steps for the Quality Measure Index and MDP TEP.

[1] Center for Clinical Standards and Quality, Centers for Medicare & Medicaid Services. *CMS Quality Measure Development Plan: Supporting the Transition to the Merit-based Incentive Payment System (MIPS) and Alternative Payment Models (APMs).* Baltimore, MD: US Department of Health and Human Services; 2016. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/Final-MDP.pdf. Accessed November 13, 2018.

III. MEETING PROCEEDINGS

Welcome and Opening Remarks

Presenter: Kyle Campbell, PharmD, HSAG

Dr. Campbell, Project Director, welcomed the TEP members and attendees from CMS. Dr. Campbell noted that HSAG was recording the meeting, provided technical guidance for participating in the webinar; and reminded participants that meeting materials are proprietary to the project and cannot be shared without permission from CMS. He displayed the *TEP Meeting Agenda* (Appendix A) and outlined the objectives of the webinar.

TEP Roll Call and Disclosures of Conflict of Interest

Presenter: Michael Phelan, MD, Cleveland Clinic Health Systems *(Co-Chair)*

Dr. Phelan conducted a roll call. One TEP member joined after the roll call, bringing the total attendance to 17 members, as indicated by the checkboxes.

☒ Peter Aran, MD
☐ Brandy Cunningham, MS
☒ Lindsay Erickson, MSPH
☒ Robert Fields, MD, MHA
☒ Eliot Fishman, PhD
☒ Jeremy Furniss, OTD, OTR/L, BCG
☒ Lisa Gall, DNP, RN, FNP, LHIT
☒ Rachel Harrington, PhD
☒ Mark Huang, MD
☐ Kent Huston, MD
☒ Joel Kaufman, MD, FAAN
☒ Erin Mackay, MPH
☐ Scott Mash, MSLIT, CPHIMS, FHIMSS
☒ Giselle Mosnaim, MD, MS, FAAAAI, FACAAI
☐ Amy Mullins, MD, CPE, FAAFP (*TEP Co-Chair*)
☒ Amy Nguyen Howell, MD, MBA, FAAFP
☒ Michael Phelan, MD *(TEP Co-Chair)*
☒ Kristin Rising, MD, MSHP, FACEP
☒ Lynn Rogut, MCRP
☐ Heather Smith, PT, MPH
☐ Lisa Gale Suter, MD
☒ Samantha Tierney, MPH
☒ Lindsey Wisham, MPA

Members disclosed or restated information about potential conflicts of interest:

- L. Gall works for Stratis Health, part of the Lake Superior Quality Innovation Network (LSQIN), which has CMS contracts.
- R. Harrington has joined the National Committee for Quality Assurance, which receives grants and contracts from CMS and other federal sources; she specified that her input at the TEP would be based on her personal opinions and experience rather than those of her employer.
- M. Huang continues to participate on the National Quality Foundation (NQF) measures feedback loop committee.

- G. Mosnaim continues to hold stock options, perform research, and participate in consulting and/or advisory activities on behalf of companies she has previously disclosed.
- S. Tierney works for PCPI, which has grants and contracts with CMS and private organizations.
- L. Wisham works for Telligen, which has a contract with CMS.

Review of MDP Activities Since February 2019 Meeting

Presenter: Kendra Hanley, MS, HSAG

Ms. Hanley notified the TEP that HSAG posted the 2019 MDP Annual Report online on May 31, and she provided a link to the document. She also said that CMS has reviewed an initial draft of updates to the MDP and is determining the timing and approach to incorporating these updates.

Quality Measure Index Development Review

Presenter: Carolyn Lockwood, MSN, RN, HSAG

Ms. Lockwood reviewed the goals and potential uses of the Quality Measure Index. The index will provide a transparent framework to assess the relative value of individual measures based on measure variables and to support CMS efforts to develop and select meaningful measures to improve patient outcomes with less burden. Once developed, the index will support stakeholders as they prioritize measures for development and continued implementation, and it will inform measure developers about measures that could become more meaningful with updates.

She then reminded the TEP how the Quality Measure Index score is generated: The index user abstracts information from publicly available sources on the scoring variables (e.g., feasibility and reliability), analyzes the information, and calculates a score of zero to 100. This score places measures into the Good, Moderate, or Needs Improvement category.

Finally, Ms. Lockwood reviewed the steps that the team has taken in developing and testing the index: performing an environmental scan to identify variables; developing the abstraction tool; conducting two phases of alpha testing; convening webinar workgroups to discuss the index's content validity and weighting; and conducting beta testing, which is now complete.

TEP Comments and Feedback

- A TEP member suggested that rather than using the Quality Measure Index score categories of Good, Moderate, and Needs Improvement, the team should consider another option such as High, Middle, and Low Performer. Ms. Lockwood thanked him for his input.

Expert Workgroup Webinar Meeting: Content Validity

Presenter: Heather Tinsley, MSPH, HSAG

Ms. Tinsley acknowledged the TEP members who participated in the two workgroups, the first of which assessed the content validity of the scoring variables in the Quality Measure Index. By establishing that these variables are valid in determining the quality of measures, the team is able to infer that the index itself is valid, she said.

Nine of the 10 workgroup members responded to a pre-assessment, in which they rated 13 scoring variables as Essential, Useful But Not Essential, or Not Necessary. Recommendations from the workgroup meeting—which nine members attended—were to:

- Regard most of the scoring variables as Essential or Useful But Not Essential for beta testing.
- Remove the *Use of Measure* variable and insert its operational definitions under *Alignment.*
- Remove the operational definition of *Alignment,* which requires measures to be present in at least one of the eight Core Quality Measures Collaborative (CQMC) Core Measure Sets, since these sets do not represent many important medical specialties, and as a result this variable could encourage greater measure gaps.
- Evaluate whether the *NQF Endorsement Status* variable adds value to the index.
- Consider stratifying the *Burden* variable as high, medium, or low. (Though the team will reevaluate this step in the future, they were unable to operationalize it during beta testing since measure burden is not currently reported in a way that supports such stratification.)

TEP Comments and Feedback

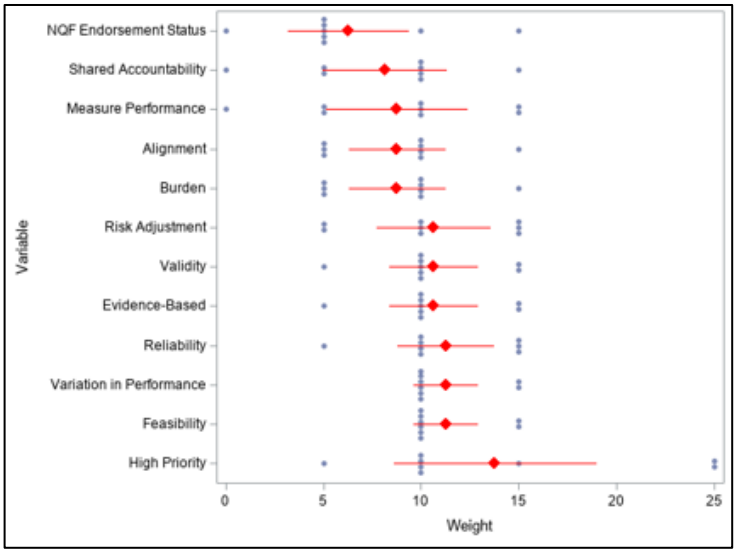
- A TEP member asked for clarification on the definition of the *Burden* variable; Ms. Tinsley explained that no methods of operationalizing cost or time burden are available, which is why the team defined *Burden* as using additional quality data codes such as CPT® Category II codes, following a recommendation from the 2016–2017 MDP TEP. She also noted that the team will consider new methods for operationalizing burden for the Quality Measure Index as they become available. The TEP member acknowledged the challenge of assessing the burden of a measure within the index.

Expert Workgroup Webinar Meeting: Weighting

Presenter: Cherrishe Brown-Bickerstaff, PhD, MPH

Dr. Brown-Bickerstaff reviewed the results of the second workgroup, which provided a forum for members to discuss alternatives to equal variable weighting, an approach in which each scoring variable has the same impact on, or contribution to, the final Quality Measure Index score. In a pre-assessment exercise, seven of the 10 members allocated a "budget" of 120 points across the 12 remaining scoring variables, based on their experience and their perceptions of the relative importance of the variables. Allocating 10 points to each variable demonstrated equal preference for all, and giving more or fewer points to a variable demonstrated a preference for the variable to have greater/lesser weight in the scoring algorithm and greater/lesser impact on index scores.

Dr. Brown-Bickerstaff displayed the results of the post-meeting survey, which repeated the exercise after members had discussed their recommendations at the meeting (right). Noting the limited sample size of post-meeting respondents (n = 8), she observed that each scoring variable received a wide range of point allocations, indicating a general lack of consensus among members.



Considerations for using an expert-based weighting scheme for scoring variables that arose from the meeting were:

- Weighting preferences are subjective based on specific stakeholders' needs.
- Evolving priorities could make maintenance of variable weighting difficult.
- Some elements of NQF endorsement and the Measure Applications Partnership (MAP) process are reflected in multiple variables (e.g., *Reliability*, *Validity*, *Feasibility*).

The workgroup recommended potentially using a weighted domain structure for grouping variables, with each variable within a domain receiving the same weight, or weighting variables with index scores displayed for each domain.

TEP Comments and Feedback

- During the presentation of the figure above, a TEP member asked whether any individual members tended to give more extreme allocations to variables (as an example, he asked whether one member gave the extremely low weight to *NQF Endorsement Status* and extremely high weight to *High Priority*). Dr. Brown-Bickerstaff responded that the team focused on the aggregate response for each scoring variable rather than tracking individual responses, and she reiterated the overall lack of consensus among members.

Quality Measure Index Beta Testing Results

Presenter: Rob Ziemba, PhD, MPH

Dr. Ziemba provided a high-level review of the beta testing results, focusing on the scoring variables' reliability, content validity, and feasibility/impact. He said that beta testing found 10 of the 12 scoring variables reliable, based on their kappa score or percent agreement between the two subject matter experts; however, the subject matter experts were unable to consistently abstract *Alignment* and *High Priority*. In terms of content validity—based on workgroup votes for variables as "Essential" or "Useful But Not Essential"— the team found that 6 of the 12 scoring variables had high content validity, and 6 had moderate content validity.

By comparing Quality Measure Index scores computed with and without each variable, the team found that 11 of 12 variables had a positive impact on index score variation (i.e., helping the

index distinguish between low- and high-quality measures), Dr. Ziemba said. The sole exception was *High Priority*. As a result of these analyses, the team removed the variables *Alignment* and *High Priority* from further testing. However, alignment and high priority are important characteristics of a quality measure and could be reintroduced into the index if further data standardization occurs.

Dr. Ziemba then provided an overview of the index score cutoffs, based on tertiles of the score distribution, for the three performance categories—Good, Moderate, and Needs Improvement, which could be renamed per the earlier TEP comment—and reviewed how calculating scores with and without *NQF Endorsement Status* affected the index score of measures. Eleven measures changed performance categories. He also said that index scores were strongly associated with NQF endorsement, making the index a complementary instrument for measure assessment along with NQF endorsement, but the index had no correlation with the ACP assessment of quality measures.

TEP Comments and Feedback

- A TEP member asked how many of the 100 measures included in beta testing were NQF endorsed; Ms. Tinsley noted 42.
- A TEP member stated that the presence of NQF endorsement in the Quality Measure Index makes these two measure evaluation approaches somewhat duplicative rather than complementary. Dr. Ziemba responded that the *NQF Endorsement Status* variable has five scoring categories (i.e.., Endorsed, Endorsed-Reserve, Not Endorsed, Endorsement Removed, and eCQM Approved for Trial Use (Not Endorsed), rather than a simple yes/no, and some measures changed performance categories when the variable was removed, demonstrating additional value in the way that the Quality Measure Index operationalizes the variable. The same TEP member asked whether this variable adds information or merely amplifies some other elements in the index; Dr. Ziemba responded that merely amplifying already captured differences would not result in measures changing performance categories that are based on the ranks of measures. Ms. Tinsley added that the *NQF Endorsement Status* variable might in the future aid in selecting between two measures with similar index scores, and Dr. Campbell said the NQF endorsement process provides a type of multi-stakeholder peer review that appears to provide additional value.

- A TEP member asked whether any of the 100 measures included in beta testing were among those that the ACP found "not valid." Dr. Ziemba said beta testing included measures that the ACP judged both valid and not valid.

The discussion moved on to variable weighting, with the observation that 10 measures changed performance categories after the team applied equal domain weights (i.e., variables are grouped into equally weighted domains with variables in each domain equally weighted) rather than equal variable weights (i.e., all variables have the same weight). The TEP members considered whether, conceptually, variables should be considered equivalent or be grouped into measure evaluation domains (NQF Endorsement, Importance, Scientific Acceptability, Feasibility, and Usability) that are equivalent.

A TEP member stated that under equal domain weighting, if a domain contains three variables, each of those variables has relatively less influence relative to a variable that is the sole constituent of its domain—an important consideration, given that most variables that the workgroup ranked relatively important are in domains where their influence on the index would be comparatively diluted. Dr. Ziemba noted that this comment highlighted the main issue for this meeting to resolve: whether the individual variables or five domains are equally important. Dr. Campbell agreed, calling the question a conceptual rather than empirical issue.

A TEP member suggested weighting each domain equally but giving variables within them different weights. Dr. Ziemba said that approach is possible, but estimating those weights and performing maintenance on them could be difficult. Another TEP member supported equal domain weighting, but was concerned that if, for example, shared accountability were ever dropped from the index due to changing expert opinions, usability would no longer influence the index score; as a result, each domain would always need a variable to represent it.

Another TEP member urged the importance of transparency in how the domains contribute to the index score for a measure; for example, in the case of a measure that has lost its NQF endorsement but scores high in other domains, giving it a high overall score. Another member supported this idea, noting that many patient-centered measures are important but may have high burden or low feasibility; outside observers will need to be able to see that tradeoff in order for the index not to "mask some of the incentives that are in play for the measurement space."

A TEP member argued against equal domain weighting, citing the prominence of NQF endorsement as an entire domain after the workgroup felt hesitant to include it as a variable. Another member agreed, stating that equal variable weighting would better represent the workgroup's perception of the importance of NQF endorsement. Another TEP member asked whether the group could consider recommending the use of unequally weighed domains. Dr. Campbell responded that the group could make that choice later in voting. Two other members agreed with this approach.

Dr. Campbell said, in response to a TEP member's question about modifying weights over time, that the domain approach would be easier to modify, since it would be easier to add emerging variables to domains without recalibrating the weights of variables in other domains. Leading into the polling, he suggested that a recommendation for domain weighting would indicate willingness to further discuss the possibility of giving domains unequal weight.

Fifteen TEP members responded, with:

- **9 recommendations for domain weighting.**
- **6 recommendations for equal variable weighting.**

As the meeting closed, Ms. Hanley stated that HSAG would continue MDP- and Quality Measure Index-related work under contract to CMS. She invited the members to convey their interest in continuing to serve on the TEP through May 2021, and she disclosed the projects slated for TEP involvement, along with their time commitment. She and Dr. Campbell thanked the TEP members for their participation and CMS for its support of the Quality Measure Index project.

APPENDIX A – TEP AGENDA

Technical Expert Panel Meeting

June 17, 2019, 12:00 p.m. to 2:00 p.m. ET

Agenda		
12:00–12:05 p.m.	**Welcome and Opening Remarks**	Kyle Campbell, PharmD HSAG
12:05–12:15 p.m.	**TEP Roll Call and Disclosures of Conflict of Interest**	Michael Phelan, MD, JD, FACEP, RDMS, CQM Cleveland Clinic Health Systems
12:15–12:20 p.m.	**Review of MDP Activities Since February 5, 2019, Webinar Meeting**	Kendra Hanley, MS HSAG
12:20–12:30 p.m.	**Quality Measure Index Development Review**	Carolyn Lockwood, MSN, RN HSAG
12:30–12:45 p.m.	**Summary of Expert Workgroup Webinar Meeting #1: Content Validity**	Heather Tinsley, MSPH HSAG
12:45–1:00 p.m.	**Summary of Expert Workgroup Webinar Meeting #2: Weighting**	Cherrishe Brown-Bickerstaff, PhD, MPH HSAG
1:00–1:50 p.m.	**Quality Measure Index Beta Testing Results**	Rob Ziemba, PhD, MPH HSAG
1:50–2:00 p.m.	**Next Steps for MDP TEP**	Kendra Hanley, MS HSAG