Public Comment Summary Report

Project Title:

Development of the Quality Measure Index (QMI)

Dates:

The Call for Public Comment ran from Friday, May 6, 2022, to Monday, June 20, 2022.

Project Overview:

The Centers for Medicare & Medicaid Services (CMS) has contracted with Health Services Advisory Group, Inc. (HSAG) to develop and refine the QMI tool. The contract name is Impact Assessment of CMS Quality and Efficiency Measures. The contract number is 75FCMC18D0026/75FCMC19F0001. As part of its development process, HSAG requested interested parties submit comments on the QMI release version 1.0 methodology. The purpose of the QMI tool is to support the assessment and selection of quality measures that provide meaningful quality performance information and align with national health care quality priorities.

Information About the Comments Received:

HSAG solicited public comments by email notification via the Measures Management System (MMS) Measure Development & Outreach Series listserv, as well as individual emails sent to relevant interested parties and entities. Additionally, a news story was posted to the MMS Hub. The call for public comment was published on the CMS MMS Public Comment website. Public comments were collected using an online comment collection tool.

We received responses from nine entities regarding the QMI tool. Specifically, from:

- Six medical or professional associations and societies
- One healthcare quality- and measurement-related organization
- Two individual consumers

We have categorized the comments into General Stakeholder Comments and Specific Stakeholder Comments, which is further delineated by these topics:

- Implementation of the QMI tool
- Scoring methodology
- Existing QMI domains and variables
- Consideration of new QMI domains or variables

Stakeholder Comments

General Stakeholder Comments:

We received comments on various aspects of the QMI tool, including questions and considerations regarding its methodology and its intended usage. Four commenters offered appreciation for the overall intent of the QMI tool.

One commenter expressed concern regarding the large number of measures across CMS quality programs and suggested narrower, more specific measure cohort definitions; these comments were determined to be out of scope for this public comment.

Specific Stakeholder Comments: Implementation of the QMI Tool

1. **<u>Purpose of the QMI tool and process for implementation</u>**

Five commenters requested clarification about the purpose of the tool and the process for implementing it.

<u>Purpose of the QMI tool</u>

- Three commenters addressed the purpose and intended use of the QMI tool.
- One commenter sought clarity about the added value the QMI tool brings to existing measure review processes.
- Three commenters cautioned that the use of the QMI tool should function as an initial screening tool in addition to expert assessment and recommendations.

Response:

We appreciate the comments. The QMI tool is being designed as a screening tool that aims to utilize a systematic and objective assessment approach, so that CMS can rapidly obtain an indication of the relative value of a quality measure. The QMI tool will continually be refined, and further evaluation of its applicability across phases of the measure lifecycle is planned.

The QMI tool is not intended to replace or negate the need and value of subsequent multi-stakeholder expert review processes conducted by CMS and the Measure Applications Partnership (MAP) or performed by the National Quality Forum (NQF) or other consensus-based entities (CBEs) when evaluating measures. As the QMI tool continues to undergo refinement, based on guidance from CMS, results from testing, and feedback received from the Technical Expert Panel (TEP) and other stakeholders via Public Comment, we will continue to work closely with CMS to determine how best to incorporate the tool into existing processes.

<u>Process for implementing the QMI tool</u>

- Three commenters requested clarification regarding the different entities involved in the QMI measure review and evaluation process and their respective roles, particularly with respect to qualitative expert review of measures.
- Two commenters suggested that the QMI measure review and evaluation process should include a feedback mechanism to inform measure developers of QMI scores and allow for developers to submit clarifications or additional data.
- One commenter asked for a mechanism to appeal QMI scores.
- One commenter remarked that the QMI tool needs further revision before it achieves its intended purpose.
- One commenter suggested that an additional public comment period be held after the QMI tool has established new scoring thresholds.
- One commenter asked for additional detail and clarity in the methodology report related to variable definitions and explanations for scoring.

Response:

We appreciate the commenters' feedback. Presently, the QMI tool is being tested primarily on new measures being submitted through the CMS Measures Under Consideration (MUC) Entry/Review Information Tool (MERIT) during pre-rulemaking. Several discrete data fields were added to MERIT 2022

to standardize measure information, so the QMI tool could evaluate the scoring algorithms. Because the tool does not determine appropriateness or accuracy of information submitted by measure stewards, additional expert review is sometimes warranted. Qualitative reviews throughout the measure evaluation process can serve to determine appropriateness and quality of information provided.

We will share the suggestion with CMS for a feedback mechanism to inform measure developers of QMI scores, so developers can be alerted of potential areas of improvement, have adequate time to address areas needing improvement, provide additional data or respond to low scoring QMI variables, and/or appeal scores that are generated.

We intend to utilize information received from public comment and to continue to engage with the TEP to explore approaches to address known limitations and to further refine the tool to better meet its intended purpose. As development of the QMI tool progresses, we will continue to prioritize objectivity and transparency in the QMI measure review and evaluation process.

2. <u>Alignment of the QMI tool with other quality measure evaluation criteria and processes</u>

Five commenters shared feedback regarding the alignment of the QMI tool with other quality measure evaluation criteria and processes.

- Four commenters requested clarification regarding the processes by which the QMI tool will align and work together with other entities and within existing measure evaluation processes that include the NQF and the MAP Measure Selection Criteria (MSC) and Measure Set Review (MSR).
- One commenter commended the QMI tool for the extent of its alignment with NQF's measure evaluation criteria and work by the NQF Scientific Methods Panel. Another commenter suggested that changes in NQF guideline recommendations should be carefully evaluated before aligning the QMI tool with them to avoid undue burden on measure developers.
- One commenter recommended when aligning the QMI tool with NQF evaluation criteria that CMS develop a process to update the QMI tool at the same time NQF modifies its evaluation criteria.
- One commenter suggested alignment with NQF's Core Quality Measure Collaborative (CQMC), specifically for the identification of disparities-sensitive measures for further health equity assessment.

Response:

We appreciate the commenters' questions and input. We agree that it is essential to distinguish the use of the QMI tool from the role of established multi-stakeholder expert review processes. The QMI tool is not intended or designed to replace expert reviews performed by NQF or other CBEs, including the MAP, when evaluating measures. Rather, the QMI tool is intended to provide a rapid and objective initial indication of the potential suitability of a quality measure for CMS programs. Presently, the QMI tool is being evaluated on measures being submitted via MERIT during the pre-rulemaking process. In the future, the QMI tool can be adapted for use in other CMS measure evaluation processes, such as with measure removal from CMS programs or to evaluate measures under development. At present, criteria within the QMI tool are developed with the aim to align, where feasible, with standards defined in the CMS MMS Blueprint and the NQF Measure Evaluation Criteria. As the development of the QMI tool continues, we will continue to vet and examine for appropriateness the latest guidance in the measurement industry prior to inclusion in the QMI tool.

We also appreciate the suggestion on how the QMI tool may explore ways to examine gaps in quality of care between disadvantaged populations and to strive to align with NQF's previously developed criteria to identify disparities-sensitive measures. We also note the ongoing work by NQF's CQMC in trying to incorporate the concept of identifying disparities-sensitive measures when selecting measures for the core sets. We aim to follow these developments made by NQF and to align to the extent feasible.

As the QMI tool's form and function evolve, we can explore alignment with the MAP MSC and MSR criteria, given the overlapping goals.

3. **Burden for measure developers**

Three commenters expressed concern regarding burden for measure developers.

- Two commenters suggested seeking harmonization and measure submission interoperability between the QMI tool and CMS/NQF forms, including CMS' Call for Measures MERIT form, the application form for Qualified Clinical Data Registry (QCDR) self-nomination, and NQF endorsement measure submission forms, so that information can be exported from one system to the other.
- Two commenters expressed concern regarding the measure testing burden on measure developers with one specifically highlighting the requirement for all critical data elements to be tested for data element validity.

Response:

We appreciate the feedback and will share the recommendation for measure submission interoperability with CMS. In 2022, to assist with development and testing of the QMI tool, several discrete fields were added to the MERIT 2022 form to explore the ability of the QMI tool to obtain adequate data in a more standardized format during the measure submission process. QMI scores can then be calculated directly from the measure information submitted by developers via MERIT.

Current industry standards including the CMS MMS Blueprint and the NQF Measure Evaluation Criteria state that if data element-level testing is conducted, testing of all critical data elements is preferred. The CMS MMS Blueprint additionally states that results for the numerator, denominator, and exclusions, at a minimum, should be provided. The NQF Measure Evaluation Criteria state that data element-level testing (now referred to as "patient/encounter-level testing") should be conducted on "ALL critical patient/encounter level elements," but that at minimum, results for the numerator, denominator, and exclusions or exceptions must be assessed and reported separately. In addition, the current CMS definition of fully developed measures, released March 1, 2022, (https://www.cms.gov/files/document/fully-developed-measure-definition.pdf-0) requires data element-level testing for each critical data element. The QMI tool has previously allowed results rolled up to the denominator and numerator level. We will consider public comment feedback related to measure developer burden, as well as NQF and CMS requirements, as we continue to evaluate requirements for data element-level testing.

4. **Inclusion of healthcare worker–reported measures**

One commenter suggested that the QMI tool also evaluate measures that assess healthcare worker–reported outcomes and experience, such as staffing levels, safety/workplace violence, and staff engagement.

Response:

We thank the commenter for the suggestion. The QMI tool is presently able to assess all measure types, as defined in the latest CMS MMS Blueprint, except for measures that are classified as structural measures or cost and resource use measures. We recognize that structural measures may capture concepts provided in the comment, such as staffing levels. CMS may explore expansion of the QMI tool to additional measure types in the future.

Specific Stakeholder Comments: Scoring Methodology

5. **Variable-level scoring**

One commenter sought clarification regarding the interpretation of QMI variable scores and whether they are discrete or continuous values.

Response:

We appreciate the commenter's questions on variable scoring. QMI scoring variables are ordinal variables characterized by four categories each. Each category is assigned one of four discrete numerical values: 0.00 (Grey – Unable to Determine), 0.25 (Red – Not Preferred), 0.75 (Yellow - Acceptable), and 1.00 (Green – Preferred). The numerical labels allow for the calculation of QMI domain and overall scores. All variables within a domain are equally weighted during computation of a domain score, and domains are equally weighted during the computation of an overall score. However, we will continue to consider and test scoring approaches as the QMI tool evolves.

6. **Calculation of domain and overall QMI scores**

Five commenters suggested that QMI variables and/or domains be weighted unequally.

- One commenter suggested weighting *High Priority* more than other categories; another suggested weighting variables that assess national strategic priorities (e.g., a future health equity variable) more than other variables.
- One commenter recommended an approach to assess the equal weighting methodology with respect to the purpose of the QMI tool.
- One commenter contended that, because some QMI variables are based on concrete data and some are based on expert judgement, it would be reasonable to weight variables differently to compensate for disparities in scoring.
- One commenter contended that the Scientific Acceptability domain should be weighted more than the other domains because it is within the measure developers' control to meet variable-specific requirements.
- One commenter disagreed with equal weighting across domains, citing that in order to be consistent with the CMS Meaningful Measures Initiative and to promote evidence-based care across CMS programs, the measure's score for the Importance domain and the Feasibility and Usability domain should carry more weight than the Scientific Acceptability domain score.

Response:

We thank the commenters for their feedback and suggestions, including the option of assigning higher weights to the *High Priority* variable; offering higher weights to specific CMS' agency-wide strategic priorities, such as when a measure addresses health equity; and a potential approach to evaluate the face validity of the QMI tool with respect to CMS' pre-identified definition of a high- or low-priority measure for inclusion in CMS programs.

Applying differential weighting of variables may be challenging since different stakeholders place higher or lower value on individual variables or domains as illustrated by the comments received. We recognize that the current method of designating QMI scores for certain variables may be based on subjective criteria that may change over time, while some variables aim to mimic more objective criteria already delineated by NQF. In addition, we appreciate the commenter's suggestion to factor in how much of a measure's attribute is perceived to be within the measure developer's control. We also acknowledge that re-weighting could occur primarily at the domain level. These are options we could consider as we explore various approaches to re-weighting and to ensure the end result meets CMS' intended purpose for the QMI tool.

We will continue to work with CMS and the TEP to explore ways of determining appropriate scoring weights and assess the implications of each approach on CMS' overall measure evaluation processes.

7. **Interpretation of overall QMI scores**

One commenter expressed concern that the QMI tool's current scoring methodology could create artificial variation in overall scores and hide significant flaws within the measures.

Response:

We appreciate the commenter's concerns that the application of an equal weighting scheme across variables and the absence of must-pass criteria may create an impression that a measure that is, for example, scientifically sound but lacks feasibility is similar to a measure that is highly feasible but lacks scientific acceptability. With all else being equal, a measure that fails *Reliability* and *Validity* could end up with the same score as a measure that fails *High Priority* and *Feasibility*. While the former may signify a significant flaw in the measure's scientific integrity, the latter may reflect a limitation of the measure from an implementation standpoint. While the distinction between a flawed measure and a measure with limitations is an important one, the QMI tool is at present being designed to support existing CMS measure review processes and not to replace them. As CMS continues to test the QMI tool and its applicability in existing measure review and evaluation processes, we will utilize feedback received during public comment and via the TEP to refine the tool and explore other ways to weight the variables.

Specific Stakeholder Comments: Existing QMI Domains and Variables

8. **Classification variables**

Three commenters made suggestions regarding QMI classification variables.

- One commenter suggested that the name of the *Measure Submission Method* classification variable be changed to "Collection Type" or "Data Source."
- One commenter suggested that an example of "Other" for the *Measure Submission Method* classification variable be provided in the operational definition.
- One commenter suggested the *Core Quality Measures Collaborative (CQMC) Measure Sets* is not a useful classification variable.

Response:

We appreciate the comments. While response options captured by the *Measure Submission Method* classification variable can be closely related to the data source, they more accurately aim to capture how the measure score would be reported to CMS. According to the MUC data template from 2021 and 2022, Clinical Quality Measures (CQMs) are considered a data submission method as captured by the

field "How is the measure expected to be reported to the program?" However, we acknowledge that in the Merit-based Incentive Payment System (MIPS) program, CQMs are classified as a Collection Type.

The response option "Other" for the *Measure Submission Method* classification variable captures descriptions that developers enter into various measure information sources, including the MERIT form, that cannot be readily identified using existing discrete answer options. Example responses for "Other" include: "NHSN" and "Submitted and linked with administrative claims data." We evaluate non-specific responses received during testing to determine how best to introduce added clarity to measure information sources such as the MERIT form in the future.

All classification variables, including the *CQMC Measure Sets* variable, are informational variables intended to identify features that do not necessarily predict the relative value of a quality measure. CQMC utilizes a multi-stakeholder expert review process and applies criteria that primarily focus on seeking measurement alignment across payers. As such, within the context of the QMI tool, the *CQMC Measure Sets* variable is intended to flag a measure that has been included in any one of the 10 core sets of quality measures and may serve as a signal for potential alignment. We will continue to work with CMS to determine the usefulness of this and other classification variables within the QMI tool as development continues.

9. <u>Importance domain: Evidence-based variable</u>

Three commenters provided feedback regarding the Evidence-based *variable, under the Importance domain.*

- One commenter emphasized the importance of the *Evidence-based* variable in the QMI tool and urged CMS not to include measures in its programs that are inconsistent with recommendations provided by national clinical guidelines.
- Two commenters requested clarification regarding the operational definitions for guidelines based on consensus recommendations or expert opinion and how these are scored.
- One commenter requested clarification regarding the operational definition for strong or moderate guideline recommendations and whether these recommendations must be evidence-based or can be based on consensus opinion.
- One commenter stated that measures with an equity-related focus may not have accompanying clinical guideline recommendations or evidence grading, so empirical data should be sufficient.
- One commenter sought clarification regarding the type of article required to meet "Outcome measures with at least one citation."
- One commenter stated that conclusions from systematic reviews can be easily interpreted and may not need to be accompanied by guideline recommendations in order to implement the principle of care so should not be penalized with a lower score of 0.75 (Yellow – Acceptable).

Response:

We appreciate the commenter's feedback about the *Evidence-based* variable, including the importance of including the variable in the QMI tool. The NQF evaluation criteria and CMS MMS Blueprint emphasize the importance of measure development being guided by evidence-based clinical guidelines, if available, and/or translating supporting empirical evidence into meaningful quality measures. The QMI tool aligns with these industry standards to the extent feasible.

The current *Evidence-based* variable for measure types other than outcome measures or patient-reported outcome performance measures (PRO-PMs) (e.g., process measures) does not examine the quality or the appropriateness of the evidence that is being cited in measure information sources but

focuses on the type of evidence that is being cited, based primarily on a standard evidence pyramid. In addition, while a clinical guideline is ranked higher than other types of evidence in the evidence pyramid, we recognize that not all clinical guidelines that claim to be evidence-based adhere to the same standards. Hence, the QMI tool in its present version does not incorporate the process by which the clinical guideline is developed. The QMI tool instead takes the strength of recommendation and evidence grading reported by the measure developer at face value. In MERIT 2022, fields were added to capture a guideline statement's strength of recommendation and level of evidence more consistently. We will explore patterns by which developers selected responses to these fields to determine how best to capture a variety of grading conventions.

Measures submitted from developers who explicitly state that the clinical guideline panel assigned the statement that supports a measure concept a specific strength of recommendation of "strong" or "moderate" are awarded a score of 1.00 (Green – Preferred). Guideline statements based on expert opinion can come from a non-evidence-based guideline or an evidence-based guideline. Evidence-based guidelines can utilize various approaches to classify recommendations. Inserting an expert opinion in an evidence-based guideline is a choice not performed by all guideline development entities and traditionally means that the statement is crafted based on the panel's experience, knowledge, and judgment for which there may not be published evidence to support the recommendation. As such, these types of recommendations are grouped into the middle category in the QMI tool and receive a score of 0.75 (Yellow – Acceptable). Similarly, measures from developers who refer to a clinical guideline as the source of evidence to support the measure, but do not provide any grading or explicitly state that the recommendations are based on expert opinion, are currently awarded a score of 0.75 (Yellow – Acceptable) as well. This approach separates statements that may be driven by a review of the body of evidence from those that may purely be based on opinions. As development of the QMI tool progresses, we will explore whether it is feasible for it to capture nuances related to the quality and trustworthiness of clinical guidelines and how that information may contribute to scoring refinements for the *Evidence-based* variable.

Although the conduct of systematic reviews involves a team of experts and is often the basis by which evidence-based guidelines are developed, there is wide variation in the approach taken to perform systematic reviews and to report synthesized information and translate findings to a recommendation. Since the QMI tool must rely on standardized data that can be consistently provided by developers, it is at present not possible to truly align with the NQF evaluation criteria and offer a grade of High, Moderate, or Low, based on the quantity, quality, and consistency (QQC) of evidence provided in a systematic review. As such, the QMI tool offers the middle range score of 0.75 (Yellow – Acceptable).

When developers provide empirical data alone (e.g., individual peer-reviewed original studies, unpublished internal data analysis, reports published by various organizations), the QMI tool awards a score of 0.25 (Red – Not Preferred), since it is not currently feasible for the QMI tool to incorporate QQC assessment. This score may serve as an informative early indicator to CMS that the empirical data provided must be carefully reviewed and translated into meaningful actions. This type of evidence, therefore, may require extra attention during subsequent steps in the measure review and evaluation process.

For outcome measures and PRO-PMs, the QMI tool aligns with the NQF evaluation criteria in that there is no specific requirement that specific types of evidence are preferred. The NQF evaluation criteria state that evidence is present when there is empirical data demonstrating a relationship between the outcome to at least one healthcare structure, process, intervention, or service. The QMI tool at present only requires that developers provide a synthesis of evidence with accompanying references to allow for secondary reviewers to assess the applicability of the body of evidence in relation to the measure being

proposed. We will offer better clarity in the operational definition in future QMI documentation.

Lastly, we acknowledge that certain measurement areas may lack clinical guideline statements that actively urge clinicians or accountable entities to perform a specific action. We will continue to assess the feasibility and implications of incorporating QQC assessment into the QMI scoring approach for the *Evidence-based* variable as we refine the QMI tool to meet its intended purpose.

10. Importance domain: *Measure Performance* variable

Four commenters expressed concern and made suggestions regarding the Measure Performance *variable under the Importance domain.*

- Two commenters voiced concern that new measures would be penalized by the current methodology because they may not yet have measure performance data. To address this, one commenter suggested using a base score as a placeholder until a benchmark could be established.
- One commenter requested clarification on the data requirements for performance data and identified issues with data sources.
- Two commenters suggested reassessing the room-for-improvement threshold of 5%, with one commenter suggesting a higher threshold and one commenter suggesting a lower threshold.
- One commenter suggested an assessment of the range of measure performance scores across accountable entities, rather than rely on a mean score to calculate room for improvement, as currently specified.

Response:

The *Measure Performance* variable aims to capture whether there is evidence of overall room for improvement. When we explored existing measure information sources to determine what aspects of the performance scores are consistently reported by developers, we found that developers most consistently reported the mean and the extreme values often without additional details for individual percentiles to allow for greater understanding of the performance score distribution. We also found that when mean scores were not reported, usually no other accompanying values were also provided.

At present, overall room for improvement is assessed by calculating the relative difference between the mean performance score and an appropriate benchmark. The term benchmark that the QMI tool applies to the *Measure Performance* variable is the extreme values (e.g., minimum or maximum), taken from the performance score distribution across accountable entities. In the case of new measures, such as those submitted through MERIT during the pre-rulemaking process, the QMI tool uses the minimum or maximum performance scores across accountable entities reported by the developer in the data field “Minimum performance score” or “Maximum performance score,” respectively, in the MERIT 2022 form. These fields are used to calculate the relative room for improvement, depending on whether a measure’s score interpretation is “Lower score is better” or "Higher score is better.” For proportion measures only, a theoretical benchmark of 0 for “Lower score is better” measures or 100 for "Higher score is better” measures is applied if the developer does not provide a minimum or maximum performance score across accountable entities from the analyzed sample.

Performance data may come from data that were used to perform reliability and validity testing of the performance measure. Or, in the case of a measure that has been implemented, the data used to demonstrate the range of performance scores across accountable entities may come directly as a result of implementation of the measure. Performance data for the specific measure should be analyzed at the

level of analysis (e.g., facility) that the measure is specified. As such, the data used during reliability and validity testing can be a useful first step to presenting performance scores.

As an initial screening tool, the *Measure Performance* variable is presently designed to indicate that a measure with a minimum 5% relative difference between the mean and the minimum or maximum performance score is unlikely to be topped out. These measures would all receive the highest score of 1.00 (Green – Preferred). It is important to note, however, that QMI scores are not intended to be used alone to make decisions about a measure's suitability for inclusion in a program but support just one part of the measure assessment and review process. A measure that does not demonstrate sufficient overall room for improvement may not be suitable for consideration in a CMS program. Additional considerations for subsequent qualitative review include determining if performance gaps are sufficiently robust and representative of the entities being measured and whether there may be disparities in care despite low overall room for improvement.

As development of the QMI tool evolves, we will examine other approaches of determining room for improvement, including assessment of various ways to apply thresholds to align closely with the intent and purpose of the tool.

11. Importance domain: *High Priority* variable

Two commenters provided feedback regarding the High Priority *variable under the Importance domain.*

- One commenter suggested that the current method for assessing high priority areas may be too narrow and does not consider program-specific needs and priorities.
- One commenter stated that flagging digital quality measures as high priority seems premature.
- One commenter expressed concern that the newest version of the Meaningful Measures framework is not being utilized.

Response:

The priorities listed in the *High Priority* variable are determined by CMS and align with current CMS agency-wide priorities. These priorities do not capture all of CMS' priorities nor do they capture unique program-level priorities.

One of the strategic priorities tracked within the *High Priority* variable is whether the measure is a digital measure. *Digital Measure* is a classification variable in the QMI tool, defined as a measure that uses data from at least one of the following sources: electronic health record (EHR), health information exchange (HIE), registry, case management system, administrative claims, electronic patient assessment data, and wearable devices. As the field evolves and key information becomes consistently available in measure information sources, the QMI tool's *Digital Measure* operational definition will be further refined.

Related to the Meaningful Measures framework, testing of the QMI tool prior to public comment used Meaningful Measures 1.0 Priorities and Areas, as this was what was incorporated into the measure information sources at the time, e.g., CMS Measure Inventory Tool (CMIT). With the official launch of Meaningful Measures 2.0, the *Meaningful Measures Classification* variable in the QMI tool has been updated to incorporate Meaningful Measures 2.0. The MERIT 2022 form contained a field called "What one healthcare domain applies to this measure?" This field included all eight of the Meaningful Measures 2.0 priorities, and the QMI tool will evaluate the information obtained through this field going forward.

12. Scientific Acceptability domain: *Reliability and Validity* variables

Four commenters provided input regarding the Reliability *and* Validity *variables, under the Scientific Acceptability domain.*

- One commenter stated that measures in early stages of development may not have testing results available, so the QMI *Reliability* and *Validity* variables may not apply.
- Two commenters expressed concern regarding the use of specific tests and thresholds for reliability and validity testing, as outlined in the QMI methodology, which the commenters highlighted have not been finalized by NQF or articulated in the MMS Blueprint.
- One commenter advised careful evaluation of changes in NQF's recommendations regarding thresholds for reliability testing to minimize undue burden on measure developers and stated rigorous tests should be considered thoroughly before inclusion in the QMI tool.
- Two commenters provided feedback regarding the scoring algorithms for reliability and validity testing, where a decision point requires that testing was performed at the same level of intended use. One commenter requested clarification about what it means to be tested at the same level of intended use, and another commenter reaffirmed the importance of ensuring that measures are tested at the intended level of analysis.
- One commenter requested the eta-squared thresholds to delineate a small, moderate, and large effect size, as referenced in the measure score reliability scoring algorithm.
- One commenter requested clarification for the survey level testing scoring logic, specifically, whether surveys that have undergone psychometric testing score "yellow," regardless of the psychometric testing results.

Response:

We appreciate the commenters' feedback on the *Reliability* and *Validity* QMI scoring variables. Presently, the criteria used in the QMI tool is most applicable to and tested on fully developed measures. As the QMI tool is adapted for other stages of the measure lifecycle, we will reevaluate which criteria are appropriate for inclusion during each development stage.

The types of tests included in the QMI tool to assess the reliability and validity of a measure align with the CMS MMS Blueprint. Given that the QMI tool aims to standardize the initial review process for measures submitted to CMS and ensure that reviewers have the information they need to accurately assess measure data submitted for review, the QMI tool is being developed with the intention of applying statistical thresholds of scientific acceptability to the recommended statistical tests as well.

We acknowledge that the thresholds provided by NQF in the draft Acceptable Reliability Thresholds (Version 3.2) document, released in July 2021, had not been finalized or incorporated into NQF's measure evaluation criteria. As of December 2021, NQF released updated thresholds via the Scientific Methods Panel (SMP) Recommended Acceptable Reliability Thresholds (Version 3.3) document, which is not yet finalized. As NQF notes, thresholds are provided only for statistical tests commonly used and reported during testing of performance measures. We will continue to follow the work of the NQF SMP as we proceed with refining and testing the QMI tool to ensure that the tool is in alignment with NQF Measure Evaluation Criteria and thresholds published by the NQF SMP to the extent feasible and that it provides meaningful information to secondary reviewers at CMS.

Due to the lack of consensus in the field on thresholds for eta-squared analyses, the QMI tool accepted the measure developer's interpretation of whether the statistical result represents a large, moderate, or small effect size in prior rounds of testing. In future versions of the QMI tool, we will reevaluate whether

to continue to accept eta-squared tests. Further standardization in testing will facilitate fairer comparison of results across measures.

The QMI tool requires that testing be provided at the same level of analysis as the intended use. For example, if a measure was previously tested for health plan use and has been re-specified and submitted for use in an inpatient program, the measure developer should submit testing at the inpatient facility level. We will provide clarification on this point in future QMI documentation.

Related to survey-level testing, a measure would score 0.75 (Yellow – Acceptable) for the *Reliability* or *Validity* variables, if the only testing provided in the reliability or validity testing section of the measure submission form was survey-level testing. The rationale for this approach was that previous versions of the NQF and CMS forms did not have separate sections for performance measure-level testing and survey-level testing, so it was not clear to developers that both types of testing were required. The current QMI algorithms for *Reliability* and *Validity* variables were intentionally designed to give the developer credit, despite there not being any performance measure-level testing, which should be provided separately. If this allowance was not incorporated into the QMI *Reliability* and *Validity* variables, measures that lack performance measure-level testing, face validity or data element-level testing received a QMI score of 0.00 (Grey – Unable to Determine) for these variables, although the developers provided some testing information in the form of survey-level testing. In 2022, CMS separated these two requirements in the MERIT submission form, which should clarify for developers that both types of testing should be provided moving forward, as applicable. We will reevaluate how best to consider survey-level testing in future versions of the QMI tool as better data from developers become available.

13. Scientific Acceptability domain: Data element testing component for *Reliability* and *Validity* variables

Two commenters requested clarification regarding the scoring algorithms for the data element reliability and validity testing components, under the Scientific Acceptability domain.

- One commenter requested clarification regarding the relationship between measure score testing and data element testing and why the highest score that can be achieved for data element level testing is 0.75 (Yellow – Acceptable) instead of a 1.00 (Green – Preferred), as with measure-score–level testing.
- One commenter sought clarification regarding the use and interpretation of the yellow scoring category, which typically carries a value of 0.75 (Yellow – Acceptable) for the Empiric Measure Score Validity component. The commenter highlighted that under the empiric measure score validity component, the yellow scoring category is labeled as Not Applicable but assigned a score of 0.75, and this could be interpreted as not having a correlation analysis being better than having poor correlation results.
- One commenter requested clarification on why the QMI tool would use current data element standards for evaluation, instead of future standards, such as those for digital quality measures.

Response:

We appreciate the commenters' questions and comments. This version of the QMI tool aligns with the NQF Measure Evaluation Criteria to the extent feasible. In NQF's current measure evaluation algorithm, if a measure developer only provides data element–level testing to demonstrate reliability or validity of a measure, the highest score the measure can receive for that criterion is "Moderate." In the QMI scoring methodology, the equivalent score is 0.75 (Yellow – Acceptable).

The rationale for this approach is that data element–level testing alone does not provide information on how the measure performs overall. For example, a measure can have reliable data elements, but the measure scores may not be reliable. Measure score reliability and validity are important factors for CMS to consider when evaluating whether to implement a measure in its programs, particularly when measure scores are to be publicly reported or used as the basis for payment decisions. For this reason, measure score level–testing receives a higher score on the QMI tool than data element–level testing. We will reevaluate the *Reliability* and *Validity* variable scoring algorithms to align with CMS measure submission criteria as they continue to evolve.

The intent of the Empiric Measure Score Validity component was to score 1.00 (Green – Preferred) for a correlation in the hypothesized direction and 0.25 (Red – Not Preferred) if no correlation was found. The Yellow classification is thus not applicable (N/A) for this component. We will update future QMI documentation to clarify this.

We will also reconsider how best to score poor reliability and validity testing results compared to missing testing results in future versions of the QMI tool.

The QMI tool is designed to assess measures using current standards and CMS measure submission requirements. As the field continues to evolve toward digital measures, the QMI data elements will be updated to align with the criteria for digital quality measurement.

14. Scientific Acceptability domain: Face validity component of the *Validity* variable

Two commenters provided input on the face validity component under the Scientific Acceptability domain.

- One commenter suggested that rigorous face validity testing should not be awarded lower points than empiric or data element–level testing.
- Two commenters requested clear guidance regarding the "adequate number of experts" required for face validity.

Response:

We appreciate the commenters' questions and suggestions. The QMI tool aligns with NQF's Measure Evaluation Criteria when possible. NQF's measure evaluation algorithm states that a measure can only receive a score of "Moderate" on validity if the measure developer provided face validity results without empiric validity testing. In the QMI tool, the equivalent score is 0.75 (Yellow – Acceptable). We will continue to review the NQF Measure Evaluation Criteria as they continue to evolve, and we will reevaluate the scoring approach presently implemented in the QMI tool in conjunction with CMS requirements, as appropriate.

We will evaluate whether there is an accepted minimum number of experts needed for a meaningful vote on the face validity of a performance measure. If identified, we will work with CMS to update guidance on conducting face validity votes with those minimum criteria.

15. Scientific Acceptability domain: *Risk Adjustment* variable

Three commenters provided feedback regarding the Risk Adjustment *variable, under the Scientific Acceptability domain.*

- Three commenters suggested that the QMI tool should evaluate the appropriateness and adequacy of a measure's risk adjustment strategy.

- One commenter requested clarification as to whether CMS is implementing risk adjustment strategies for outcome measures currently in reporting programs and requested data that reflect this new standard.

Response:

We agree that it is important to assess the adequacy of risk adjustment models to ensure they are robust and appropriately specified for the measure concept under consideration. While the QMI tool cannot replace the thorough risk adjustment methodology review process conducted by experts, we will consider whether future versions of the *Risk Adjustment* variable can be enhanced in alignment with the updated guidance from NQF on risk adjustment.

Several CMS programs currently include risk-adjusted outcome measures. To review examples of risk adjustment methodology for different types of measures that have been reported by CMS over the past decade, visit CMS.gov at the following link to download measure methodology reports: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.

16. Feasibility and Usability domain: *Feasibility* and *Provider Burden* variables

Five commenters provided feedback on the Feasibility & Usability domain, with most of the focus on the Provider Burden *variable.*

- Four commenters stated their support for the inclusion of a variable assessing a measure's burden on providers.
- One commenter suggested that the criterion used to assess provider burden should be described as the "least burdensome method is acceptably streamlined" instead of "measure uses least burdensome method available." Another commenter suggested the *Provider Burden* variable prioritize the least burdensome data source for quality measurement (e.g., claims).
- One commenter requested clarification about why the QMI scoring algorithm for *Provider Burden* awards claims-based measures with the highest score if CMS has stated that claims-based measures are being phased out. The commenter noted that if the highest score is only being awarded to administrative claims-based measures, this qualifier should be explicitly stated in the document.
- One commenter requested clarification on why the QMI's scoring algorithm considers registries as somewhat burdensome.
- One commenter expressed concern that provider burden as a construct is subjective and depends on a measure developer's opinion and their lived experience. In addition, if the QMI variable scoring approach uses expert judgment to score the current variable, then QMI scores for this variable could be biased.
- One commenter asked that consideration be given to measure reporting burden that providers with solo or small practices would experience, given that they have fewer resources/capabilities and would benefit from a simple and user-friendly data collection and reporting system.
- One commenter suggested that the assessment of provider burden should specifically focus on providers' documentation burden.
- One commenter suggested incorporating additional dimensions to provider burden, such as measure complexity, lack of harmonization of key elements for similar measures, and unintended consequences.
- One commenter suggested that the QMI *Provider Burden* variable does not include other digital quality measure data sources (e.g., wearable technology, HIEs, etc.) in the scoring algorithm.

- One commenter requested clarification about what "Some data elements are in electronically defined fields" means for the *Feasibility* variable.

Response:

We thank the commenters for their feedback related to the *Provider Burden* and *Feasibility* variables.

The *Provider Burden* variable aims to primarily address one of CMS' quality measurement strategic objectives, which is to minimize burden for providers.

We acknowledge that a variable aiming to assess the relative burden a measure poses on providers during implementation would benefit from a multi-dimensional assessment approach. The current version of the *Provider Burden* scoring variable assesses provider burden based on one objective dimension, primarily determined by the method by which measure scores are calculated. This is in part due to the lack of consistent availability and standardization of data in existing measure information sources that could be used to assess additional dimensions of provider burden. Since the QMI tool is one part of a series of measure review and evaluation processes and emphasizes the use of a systematic and transparent approach, the QMI tool employs criteria that can be objectively categorized.

By focusing on how the measure score is calculated, the current *Provider Burden* variable assigns all electronic clinical quality measures (eCQMs) and all administrative claims-based measures a score of 1.00 (Green – Preferred). Part B claims-based measures, however, require that providers obtain quality data codes (e.g., G codes) for the measure. This additional step renders Part B claims-based measures more burdensome than administrative claims-based measures; hence, they receive a lower score of 0.75 (Yellow – Acceptable) for the *Provider Burden* scoring variable. In addition, registry-based measures may encompass a variety of approaches to data acquisition and curation, measure calculation, and performance measure score reporting, but existing measure information sources lacked data of sufficient granularity to allow for exploration and testing of key aspects of burden related to the reporting of registry-based measures. The QMI tool currently assumes that the majority of registry-based measures require the use of a third-party vendor, which somewhat adds burden for providers, compared to administrative claims-based measures or eCQMs. Only manually abstracted measures receive a score of 0.25 (Red – Not Preferred), since this is considered most burdensome. Whenever measure developers reported specifying a hybrid measure, the QMI tool used the least burdensome calculation method to determine the QMI score for this variable.

During previous rounds of testing, the QMI tool tested this scoring approach using a combination of readily available data fields within various existing measure information sources because there was no specific field to clearly indicate the method of measure calculation. During MERIT 2022, a new discrete field called "Method of Measure Calculation" was added to encourage greater standardization in data that the QMI tool could use to operationalize the existing *Provider Burden* variable without requiring expert review. With the availability of more standardized data to reflect burden on providers based on the measure's calculation method, we will explore ways to potentially differentiate lower burden hybrid measures from higher burden hybrid measures as well.

We also recognize that practices with fewer resources may find it challenging to incorporate and utilize complex reporting systems. We also acknowledge the ongoing efforts to transform the quality measurement landscape to digital status. As development of the QMI tool continues, we will explore how best to incorporate other dimensions of provider burden into the *Provider Burden* scoring variable definition, including measure complexity, harmonization, and unintended consequences, for greater alignment with the NQF Measure Evaluation Criteria. In addition, we will also explore how to consistently capture data that reflect the cost involved to engage with third party vendors or the resources involved to set up and maintain a system to support eCQM or registry reporting, especially as

the field moves toward digitalization. As CMS guidance evolves, the QMI tool will explore the potential for refining the *Provider Burden* variable scoring methodology to account for additional criteria that may better reflect burden on providers.

The QMI *Feasibility* scoring variable uses the "Feasibility of Data Elements" field within the CMS MERIT or criterion 3b.1 in the NQF main measure submission form and does not reconcile these findings with the data sources reported. These fields are focused on feasibility of collecting data at the data element level and not the feasibility of executing measure code packages during implementation. When data elements needed to calculate the measure are not all available in discrete and electronically defined fields, this would meet the criterion for "Some data elements are in defined fields in electronic sources."

Specific Stakeholder Comments: Consideration of New QMI Domains or Variables

17. **Health equity**

Five commenters made suggestions regarding how to assess health equity in future iterations of the QMI.

- One commenter suggested adding variables or domains to the QMI tool specifically to capture health equity. A second commenter recommended removing health equity from the *High Priority* variable under the Importance domain and evaluating it separately. One of these commenters further recommended weighting a health equity variable/domain score more than other variables/domains.
- One commenter asked that the request for information (RFI) from CMS on risk stratification be taken into account; another suggested incorporating assessment for risk stratification when determining if a measure addresses health equity objectives.
- One commenter suggested incorporating health equity as a scoring variable to assess whether the measure concept or how the measure is constructed leads to improved equity in care or measurement.
- One commenter suggested that health disparities could be assessed within the *Evidence-based* variable, if available in the literature, or the *Measure Performance* variable.
- One commenter shared that NQF's CQMC is considering approaches to evaluate the concept of health disparities-sensitive measures to incorporate a health equity lens to its measure evaluation process.
- Three commenters expressed concern about a lack of health equity data, including patient-reported data, collected in a standardized, low-burden manner, and demographic and socioeconomic information. One commenter specifically noted a need for data for BIPOC and LGBTQ+ individuals.

Response:

We appreciate the suggestions provided by each commenter. We recognize that CMS' strategic priorities, which include health equity, are unique to current agency-wide needs and goals and can shift over time. We will explore other approaches of evaluating priorities independently of the existing Importance domain and consider the potential role of the QMI tool to offer higher weights to CMS' agency-wide strategic priorities, such as when a measure addresses health equity.

CMS sought input through the Inpatient Prospective Payment Systems Proposed Rule, published in May 2022, on key principles for measuring health care quality and disparities across multiple social risk factors and demographic variables to support efforts to advance health equity, including the use of risk stratification. The final rule was released on August 10, 2022, and we will be considering ways to

standardize expectations on how the QMI tool might account for demographic and social risk factors and their consequent intersectional implications in quality measures. We will also explore alignment with other entities, such as the CQMC, in terms of the approach it is taking to identify disparities-sensitive measures aimed at equitably improving quality of care for all.

The current lack of patient-reported data and the level of granularity associated with critical pieces of information, such as race, ethnicity, or linguistic minority groups, and other social determinants of health, may hinder the extent to which measures can properly assess and account for equity-related goals.

As we seek alignment with evolving CMS guidance and other entities, we note that the groundwork to support the development of measures that promote equitable care and outcomes is ongoing.

18. <u>Other new variables or domains</u>

Two commenters provided suggestions for potential new variables or domains.

- Two commenters suggested the addition of a variable to assess concepts, such as measure alignment, harmonization, and attribution, such that "paired" measures and measures that can be used across multiple payers and programs score higher.
- One commenter suggested the addition of variables to assess the value of a measure from the perspective of four distinct groups: clinicians, patients, purchasers, and payers.
- One commenter suggested the addition of a variable to assess whether the measure incorporates the perspective, experiences, and input of affected patients.
- One commenter voiced concern about adding an "avoided costs" variable, as reducing the cost of care should never take precedence over whether a measure is meaningful, feasible, and evidence based.

Response:

We appreciate the suggestions. In a future version of the QMI tool, we will evaluate the potential for additional variables that assess paired measures and harmonization.

The *Alignment* scoring variable was considered in prior beta testing. Additionally, attribution was previously tested as part of the *Shared Accountability* scoring variable, and the operational definition assessed the number of different levels of analysis at which the measure is specified for use. Both variables were removed in prior beta testing due to not meeting reliability criteria because of lack of having a standardized and discrete field for data collection for scoring the variable.

Regarding a performance measure's value to clinicians, patients, purchasers, and payers, the QMI tool does not currently include this information because it has not been collected in a standardized way. However, the 2022 MERIT form included questions related to whether clinicians agreed that the performance measure was actionable for them to improve quality of care. The 2022 MERIT form also included questions related to whether patients would find the performance measure helpful to inform their care, and for measures that include structured survey or assessment data, whether patients find those assessments helpful for better understanding their condition or treatment options. We will continue to explore the feasibility of incorporating this information into future versions of the QMI tool as developers begin to provide these data in more structured fields. We agree that measures should be assessed based on their ability to improve care rather than their potential for cost savings alone. We will continue to evaluate whether there is a feasible and appropriate way to consider a measure's impact on healthcare costs relative to its benefits to patients.

Preliminary Recommendations

We will explore incorporating feedback and suggestions from commenters into the QMI tool methodology. Areas of consideration include the following:

- Continue to align the QMI tool methodology with evolving industry standards produced by NQF, such as the NQF evaluation criteria and the draft SMP Acceptable Thresholds for Reliability Testing and the CMS MMS Blueprint, where feasible.
- Track the evolution of other existing measure review and evaluation processes (e.g., CQMC, MAP MSC, MAP MSR) and align, where appropriate, recognizing the unique role that each process has.
- Enhance the QMI tool's ability to evaluate the relative value of quality measures for CMS programs by either adding new variables or modifying existing variables. New variables include those that assess national strategic priorities like health equity and alignment across payers and quality programs. The modification of existing variables could include refining the scoring algorithm for *Provider Burden* to account for digital measures, electronic data sources, measure complexity, harmonization, and unintended consequences.
- Explore approaches to assigning scores at the variable level and test various variable- and/or domain-level re-weighting schemas to calculate QMI scores that may provide more meaningful results for CMS.
- Improve standardization of data collection through measure information sources, such as the MERIT form, to obtain objective data for use by the QMI tool.
- Explore feasibility and approaches to allow for a feedback loop between measure developers and CMS on QMI scores.

Overall Analysis of the Comments and Recommendations

Public comment on the methodology of the QMI tool elicited thoughtful comments from interested parties on the purpose and intent of the QMI tool as well as the scoring methodology. Commenters identified technical issues and concerns related to the construct and execution of the QMI tool's calculation of a score to reflect the relative value of a quality measure for CMS programs. We will share the feedback obtained during public comment with CMS and explore incorporating the suggestions during continued development of the QMI tool, where appropriate, to ensure the tool maintains its intended purpose as an initial screening tool.

Public Comment Verbatim Report

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
1	June 1, 2022	Erin Lee Measure Development Manager American Academy of Neurology	Nonprofit	**Q1. The QMI can be adapted over time as strategic priorities change and measurement science evolves. Please provide recommendations to refine existing variables and their operational definitions included in the QMI, and any additional variables or domains CMS should consider for future inclusion in the QMI.** The American Academy of Neurology (AAN) appreciates the opportunity to comment on the proposed Quality Measure Index (QMI) methodology report and supports the intent to assess and select quality measures that provide meaningful quality performance information. The AAN believes that a transparent and reliable scoring instrument based on standardized definitions of quantifiable measure characteristics would enhance the existing endorsement and measure select processes. The AAN has concerns with the current proposal as application of the standards would result in substantial winnowing of measures in the program, particularly specialty set measures. Variable-Level Scoring The AAN would like to see greater clarity provided in the scoring interpretation. Are the scoring values concrete or are they meant to represent a range? Would the range of 0.01-0.25 be valued as 'not preferred'? Would the range of 0.26-0.99 be valued as 'acceptable'? Could a measure score 0.01-0.025 in one category and still be approved for use in CMS programs if the other categories scored high? The AAN would recommend a feedback mechanism to notify the measure developers or stewards when certain variables receive values of 0.00. This would allow time to obtain or explain the missing piece of data that resulted in a 0.00 score. Variable-Level Scoring: Evidence-Based The AAN supports the inclusion of this variable and suggests a closer look at the language used for "Strong or moderate guideline recommendation." It is not clear from this statement if the guideline needs to be evidence-based or if it can be consensus opinion. Additionally, not all evidence-based guidelines are created equal and apply the same rigor to their methodology. The AAN would also urge reconsideration of placing "Systematic review without guideline recommendation" in the yellow tier of scoring. It is easy to interpret conclusions from systematic reviews and isn't always necessary to have a corresponding guideline recommendation in order to implement the principle of care. The AAN suggests further clarification of "Outcome measures with at least one citation." There are no requirements stipulated for the type of article required and

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	leaves a lot of room for interpretation compared to the other requirements given. Lastly, the AAN would appreciate further clarification of the requirements for "Guideline based on expert opinion." Unlike an evidence-based guideline which is required to have a strong or moderate recommendation, no guidance is provided for the strength of recommendation needed for this type of evidence. Variable-Level Scoring: High Priority The AAN supports the inclusion of a high priority variable. However, it is not clear who decides which priorities apply to a measure. The current wording stipulates that measures can only be deemed high priority if they are digital measures, outcome measures, or address health equity based on the Meaningful Measures 1.0 framework. AAN is concerned that the newest version of Meaningful Measures is not being utilized, and that one of the eight domains has been deemed more important than others. Health equity or disparity data is hard to obtain and generalize due to the lack of standardization across EHR vendors and data sets. The transition to digital quality measures is advancing, however, the transition is meant to be slow to ensure all measure developers have the resources and tools to contribute to this advancement. Indicating digital quality measures as high priority seems premature as most EHR vendors and measure developers are still exploring this area and what it means in terms of reliability, validity, and feasibility testing. Additionally, this variable is the only place where digital quality measures are mentioned which leads to confusion. Variable-Level Scoring: Measure Performance The AAN understands the need to include measure performance as a variable. However, the way this variable is currently structured, it will inevitably penalize the use of new measures until a benchmark can be established. The AAN recommends the use of a base score that is awarded to new measures until a benchmark can be determined. The AAN would appreciate clarification on the data requirements for this variable. If measure stewards are expected to provide their own performance data, the cost will be prohibitive to measure developers. Conversely, if measure stewards are expected to use CMS data, the public would need access to this data and information on how deciles are determined. Additionally, if utilizing CMS data, there is a concern that the data is not reflective of true practice gaps. CMS data is incentivized and as a result lower performing individuals do not submit data. Therefore, there may be a gap for a measure to address that is not evident in CMS QPP data. Finally, the scoring provided for this measure indicates that it's either pass or fail. There is no middle ground. The AAN hopes this scoring methodology can be revisited and find leniency for measures that don't have more than 5% room for improvement – like those at 4.9%.

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	**Q2. For the Reliability and Validity variables, please recommend additional tests or thresholds the QMI should consider acceptable.** Variable-Level Scoring: Measure Testing - Reliability, Validity The AAN believes measure testing should play a role in this framework. However, it appears there is an expectation that reliability and validity testing have been conducted utilizing the specific tests listed in this document. For measure stewards, these testing requirements have not been articulated in the CMS Blueprint or the NQF Testing Guidance which means this is a new standard for testing. The guidance used for development of reliability is a draft NQF Scientific Methods Panel and should not be relied upon until a final report is available. It is unreasonable to include this variable based on these requirements without advance notice. Currently, it costs approximately $35,000-$75,000 to test one individual measure and the time to generate needed testing contracts, data, and final reports is approximately 18-24 months. The QMI should not be rolled out until such time has been given for measure stewards to secure funds, execute testing contracts, gather testing data, and have data at the requested level. Variable-Level Scoring: Reliability (Data Element Component) The scoring table for this variable is confusing and the AAN would appreciate clarification. The highest score that can be achieved, 0.75, is indicated as 'yellow' rather than 'green', which departs from the methodology used throughout the rest of the document. The AAN is unsure why 'Not applicable' is considered green in this variable. Additionally, data elements are dependent on what is available in electronic health record fields. Given the challenges with interoperability due to unstructured data and lack of standardized EHRs, it's unclear why the QMI report chose to calculate the data component variable for current measures standards (e.g., CMS Blueprint v17.0 and the 2021 NQF measure evaluation criteria) when there are plans to move to dQMs. Digital measures will have to adhere to specific standardized data elements (USCDI), which could improve feasibility, validity, and reliability but are not referenced in the report. Shouldn't efforts to harmonize CMS measures align with guidelines for future standards (dQMs)? **Q3. Please provide feedback on the proposed scoring methodology, including the appropriateness of equal weighting of domains in calculating the overall QMI score.** Domain-Level and Overall Scoring The AAN realizes the need for simplicity in overall scoring methodology. However, some variables included in this report are based on expert judgement, while others are

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	based on concrete data. It would be reasonable, and expected, that some variables are weighted more than others to compensate for the disparities in scoring. **Q4. Since CMS has established health equity as a strategic priority, please comment on how the QMI might further incorporate health equity or disparities as a scoring variable.** Commenter skipped this question. **Q5. To reduce burden on measure developers, CMS aims to closely align the QMI with National Quality Forum (NQF) evaluation criteria. Please provide recommendations to further align the QMI with NQF evaluation criteria.** Commenter skipped this question. **Q6. Please provide general recommendations for future considerations for the QMI.** Variable-Level Scoring: Risk Adjustment The AAN believes that risk adjustment is important for accurate performance data across practices and providers. The AAN has not heard about CMS implementing risk adjustment strategies for current outcome measures in accountability programs and measure developers would be interested in seeing data that reflects this new standard. Additionally, the AAN believes any suggested risk adjustment strategy should be evaluated to determine if it is appropriate for the measure concept. There could be significant issues if an inappropriate risk adjustment strategy is implemented and linked to payment for providers. Variable-Level Scoring: Feasibility The AAN reiterates the commitment to measure testing to inform measure acceptance in accountability programs. Given the lack of standardization of data elements, how will this be defined? Will this include electronic claims form data? Is the availability of one LOINC or CPT code, or one or more value sets sufficient to meet the 'some data elements are in electronically defined fields'? The AAN would appreciate greater clarity and definition for each of these scoring categories. Again, the AAN noted there is no mention of USCDI data elements for dQMs. Variable-Level Scoring: Provider Burden The AAN welcomes the attention given to provider burden as this is a huge issue across all of healthcare. The QMI methodology aims to align with the 2021 NQF measure evaluation criteria and CMS Blueprint v17, which means dQMs and FHIR implementation were not considered in this framework. The AAN understands the transition to dQMs is to minimize physician and practice burden by improving interoperability and standardization. Digital measures are not designed for manual

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	abstracts and will include data from other sources including wearable technology, HIEs, and claims. This methodology does not provide scoring for all dQM data sources. The AAN would appreciate more information about why registries are considered somewhat burdensome if they extract the data directly from the EHRs and curate the data post-extraction. It would seem logical to separate registry into different data submission types to accurately classify the burden each of these variations pose. Many registries utilize data extraction with minimal to no burden for providers. These registries are not more burdensome than claims or eCQMs. In addition, provider burden is directly related to how their EMR system submits data through claims, eCQMs, dQMs, and registries, and not all are equal. Last, the AAN knows from research that the more components a measure contains, the more complex and burdensome it becomes. CMS routinely tells the AAN to combine measures, sometimes inappropriately, therefore increasing the burden of using and reporting on the measure. The AAN notes that provider burden is very subjective depending on the individual's opinion and their lived experience. It is indicated that this variable will require some expert judgment which will result in highly subjective, and biased, ratings. Scientific Acceptability Domain The AAN appreciates the flow charts provided in the appendices. The AAN would like to better understand if measure score testing is preferred over data element testing for reliability and validity, as implied in this methodology. Additionally, further definitions would be helpful to explain "Testing is at the same level as intended use" in the Reliability flow diagram. The AAN requests definitions for eta-squared values that are considered 'large', 'moderate', and 'small'. Is it correct to interpret that if survey level testing has undergone psychometric testing, it will automatically be categorized as 'yellow' no matter what the results of the testing are? In closing, the AAN believes this document could provide better definitions and explanations for the variables and scoring included in this methodology to offer better transparency. In addition to the information included in this document, the AAN suggests including information on who will be expected to review and score measures based on these criteria and how to refute or request a second review. Furthermore, will this framework be implemented retroactively for existing measures, or only new measures going forward, and will this review need to be redone annually? The AAN believes implementing this will have a large impact on specialty societies and their

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	ability to manage costs on a limited budget and increase staff workload and burden going forward. The AAN thanks CMS for creating this initial draft of a scoring instrument for measures in accountability programs.
2	June 3, 2022	Neha Agrawal, MPH American Society of Clinical Oncology	Medical Society	**Q1. The QMI can be adapted over time as strategic priorities change and measurement science evolves. Please provide recommendations to refine existing variables and their operational definitions included in the QMI, and any additional variables or domains CMS should consider for future inclusion in the QMI.** ASCO Comments: Table 1 Comments Additional data variables to consider would be 'paired measures,' and if any part of the measure is or will be harmonized with another measure. For the definition of measure submission method, we believe CQM designation speaks more to the data source rather than a submission method, we therefore recommend that variable name be changed to "Collection Type" or "Data Source." **Q2. For the Reliability and Validity variables, please recommend additional tests or thresholds the QMI should consider acceptable.** ASCO Comments: Under data element validity component, we wish to emphasize how infeasible/challenging it would be for medical specialty societies like us to conduct data element validity for each individual data element. Additionally, there is little value in trying to assess data element validity at such a detailed granularity since the data elements with poor validity will undoubtedly show in numerator, denominator, exclusions, and exceptions validity calculations. For face validity, it would be helpful if there were explicit guidance for the "adequate number of experts" noted in the limitations. **Q3. Please provide feedback on the proposed scoring methodology, including the appropriateness of equal weighting of domains in calculating the overall QMI score.** ASCO Comments: In table 4 under evidence, it's unclear how CMS will score "consensus" recommendations. For example, in NCCN guideline recommendations, NCCN's grading mentions consensus (outlined below), that might be interpreted as a "guideline based on expert opinion" and receive a 0.75 score:

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	Category 1: Based upon high-level evidence, there is uniform NCCN consensus that the intervention is appropriate; Category 2A: Based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate; Category 2B: Based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate; Category 3: Based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate. The limitation acknowledged for the evidence-based variable approach is an important one. We have some concerns about CMS/NQF staff's ability to operationalize. How will the secondary review of a measure's evidence base be scored if the evidence is weaker, but directly supports the measure focus? For the measure performance scoring methodology, the document states, "all measure types that do not provide a mean of accountable entity performance receive the lowest score for this variable because the relative room for improvement is unknown." We wonder if it would be more appropriate to look at the range of accountable entity performance? For validity, empiric measure score component received 0.75 for not applicable designation, while a not applicable designation for data element component validity is not scored. ASCO seeks additional clarification - Perhaps the intent was to either give a score of 1.00 for a meaningful correlation, 0.25 for no correlation, and completely skip assigning 0.75. We have concerns that poor correlation is still preferable over not even attempting to correlate with a similar measure/outcome. **Q4. Since CMS has established health equity as a strategic priority, please comment on how the QMI might further incorporate health equity or disparities as a scoring variable.** ASCO Comments: Consider that equity-related measures may not have evidence grades / strength of recommendations associated with them, in which case an evidence summary could be sufficient. Disparities could be categorized under the Importance: Measure Performance category, if available in the performance data. Disparities could also be categorized under the Importance: Evidence-based category, if available in the literature.

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	**Q5. To reduce burden on measure developers, CMS aims to closely align the QMI with National Quality Forum (NQF) evaluation criteria. Please provide recommendations to further align the QMI with NQF evaluation criteria.** ASCO Comments: We seek clarity on how this and the NQF MAP processes will work together for measure evaluation and selection for program use. Under the limitations section (3rd bullet), the document says that using the single best result can create an upward bias of QMI scores, and in these instances, supplemental qualitative expert review is required. It would be helpful to clarify who would be identifying these instances where a QMI score is perceived to be overly inflated. The information in the QMI tool is very duplicative of the information submitted in the Call for Measures Merit Form. It would be very useful to harmonize and have a system in the future that can export the information into the QMI tool to reduce developer/staff burden and resources. **Q6. Please provide general recommendations for future considerations for the QMI.** Commenter skipped this question.
3	June 3, 2022	Colleen T. Skau, PhD Assistant Director, Quality and Performance Measurement College of American Pathologists	Medical Specialty Society	**Q1. The QMI can be adapted over time as strategic priorities change and measurement science evolves. Please provide recommendations to refine existing variables and their operational definitions included in the QMI, and any additional variables or domains CMS should consider for future inclusion in the QMI.** Many of the existing variables in the QMI align with best practices as established by NQF and included in the MMS Blueprint. However, due to the intended utility of the QMI, some may not be appropriate for inclusion or for equal weighting. First, although the classification variables are not scored, they represent an important method of grouping measures and therefore should be considered carefully. We suggest that CQMC Measure Sets is not a useful classification variable. The Core Measure Sets were defined several years ago and have not been revised to align with Meaningful Measures 2.0. Furthermore, the Core Measure Sets are very limited, and the fact that a measure is in a clinical area not represented in these sets should not be taken to indicate that the measure is of lower quality or priority. Within the scoring variables, there are a few which should be given additional consideration. First, the presence of Measure Performance, while important, limits the utility of the QMI. If this tool is intended to evaluate measures for acceptance into quality improvement programs, Measure Performance may not be available at the

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	time of initial submission. Therefore, this variable as defined would not be useful. If there were a broader way to consider performance gap, it may have greater utility. Additionally, the definition of the High Priority variable is very limited and does not consider program-specific needs and priorities. Some programs may have gaps in measures that fall outside of these narrow priorities. Furthermore, some specialties may not have the ability to create or access to data necessary for outcome or PRO-PM measures. Finally, the Provider Burden variable does not appear to align with CMS's stated priorities for future measure development. CMS has indicated that claims-based measures are being phased out and indeed is no longer accepting new claims measures. However, per the QMI, a measure based entirely on claims data would receive the highest score for Provider Burden. While it is true that this is low burden for providers, other considerations mean that claims are not the ideal data source. If this is intended to account for administrative-claims-based measures such as population health measures, that should be explicitly noted. Otherwise giving full points for claims-based measures while stating that claims-based measures are being phased out sends mixed messages to measure developers. **Q2. For the Reliability and Validity variables, please recommend additional tests or thresholds the QMI should consider acceptable.** The QMI is in alignment with the recommendations of NQF regarding acceptable thresholds for Reliability. However, CMS should carefully evaluate any changes to NQF recommendations to ensure they will not pose an undue burden on measure developers. Given that a stated use of the QMI is to inform decisions about when to discontinue development, increasing thresholds or more rigorous tests should be considered thoroughly before inclusion in the QMI even if suggested by NQF. With respect to Validity, we propose that face validity should not be automatically scored lower than empiric or data element level testing. NQF accepts face validity testing for initial endorsement of a measure, as does CMS for QCDR measures. Therefore, thorough, rigorous face validity testing should be granted the same chance to achieve a full score. Furthermore it would be helpful to have explicit guidance for an acceptable number of experts to satisfy face validity testing. **Q3. Please provide feedback on the proposed scoring methodology, including the appropriateness of equal weighting of domains in calculating the overall QMI score.** We have concerns that the scoring methodology could promote artificial variation in overall scores that is not meaningful when comparing measures to each other or to a benchmark. This is due in part to the fact that scores are discrete not continuous and

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	in part to the fact that each variable has its own cutoff. For example, a measure that received a zero for reliability (a Must Pass criterion for NQF) but the highest score for all other variables could get an overall score of 89. That would exceed the score of a measure that was "acceptable" across every variable including reliability. This would hide the significant flaw of the former measure as opposed to the latter. The limitations of the QMI acknowledge that expert review may be necessary in some areas; additional information about how this would be operationalized will be necessary to avoid undoing any standardization benefits gained from using the QMI. Weighting all categories equally seems inherently unfair since a measure may be limited in some areas outside the control of a measure developer or clinicians. For instance, not every measure will meet at least two of the high priority areas. While these may be high priorities they are not the only priorities and important measures in other areas should not be disadvantaged simply because they are not PRO-PMs or are not judged to directly affect health equity. Given that the Acceptability domain is significantly more work than the other domains and it is entirely within the developers' control such that every measure could satisfy all of the variables (unlike the Importance domain), it should be weighted more heavily than Importance. This is more consistent with NQF's current scoring strategy, which rates Scientific Acceptability as "Must Pass" in contrast to some other variables such as Feasibility. **Q4. Since CMS has established health equity as a strategic priority, please comment on how the QMI might further incorporate health equity or disparities as a scoring variable.** It is not clear how the variables and scoring included in the QMI will align with CMS's recent request for information regarding efforts to stratify measures by demographics and/or social risk factors, published in the Inpatient Prospective Payment System Proposed Rule in April of 2022. We applaud CMS's efforts to risk stratify measures as risk stratification is an important step towards health equity and CMS is in the best position to do so in a thorough, consistent, and fair manner across measures and programs. However, as this is still in the RFI stages, it is not accounted for in the QMI. We suggest future iterations of the QMI take this into account before mandating risk stratification by developers. Until significantly more patient-reported data can be collected in a standardized, low-burden manner, it is difficult to see how additional variables around health equity or disparities can reasonably be included in the QMI. Patient-reported data is the gold standard for any equity efforts, and CMS should focus energy on incentivizing ways to collect it without penalizing clinicians (such as non-patient-facing clinicians) who cannot do so.

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	**Q5. To reduce burden on measure developers, CMS aims to closely align the QMI with National Quality Forum (NQF) evaluation criteria. Please provide recommendations to further align the QMI with NQF evaluation criteria.** As noted above, NQF allows thorough, rigorous face validity testing for initial endorsement. Allowing face validity to receive full scoring would reduce burden on measure developers and align with NQF evaluation criteria. CMS should also consider whether there is a common format that would allow measure developers to input information once and have it used for both submission to CMS MERIT, QCDR self-nomination, and submission to NQF for endorsement. Even if the systems cannot connect directly to each other, an exported file that can be uploaded into either system would be a step forward; in short, measure submission interoperability. **Q6. Please provide general recommendations for future considerations for the QMI.** While we applaud CMS's efforts to standardize measure evaluation within and across programs, a tool such as the QMI must be carefully constructed and deployed to ensure that it does not unnecessarily increase burden on developers or worse, disadvantage certain types of valuable measures and specialties because they do not fit within the narrow confines of the tool. In order to guarantee fair and accurate use of the tool, scores and feedback should be shared with measure developers before they are finalized, and developers should be allowed to comment or provide additional data for low-scoring variables. In fact, we suggest that, since the output of the QMI is a single number, it should be used as an initial screening tool and the final assessment conducted by experts. Although automation of measure scoring may be appealing, nuanced judgements of the evidence, importance, and quality of a measure will still be necessary. Furthermore, in order to warrant use of the QMI and any accompanying changes to the development or testing process, increased value to measure development must be demonstrated. That is, CMS should show not just that the tool can used to evaluate measures but that doing so creates better measures, lowers burden on clinicians and/or measure developers, or improves outcomes. Use of the QMI should demonstrably enhance some aspect of the measure development or approval process, and feedback provided to measure developers should be meaningful above and beyond what is already provided. In fact, CMS should be explicit about how and when the tool will be used and by whom. The QMI states that it is intended to enhance not replace existing processes. However, current measure approval processes such as submission to MERIT for inclusion on the MUC list are already labor-intensive and time-consuming, and evaluation takes months. Additional specificity about how this

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	tool will streamline or improve this process will be critical, perhaps even description of specific use cases. Furthermore, CMS should be explicit about whether and where program-specific needs will be taken into consideration or flexibility allowed based on competing priorities. We do believe that if implemented in a transparent and fair fashion with appropriate expert input, the QMI could help facilitate creation of meaningful quality payment measures to improve patient care across CSM programs and the care continuum.
4	June 5, 2022	Denise Johnson, LCSW-C Senior Practice Associate National Association of Social Workers	Respondent skipped this question	**Q1. The QMI can be adapted over time as strategic priorities change and measurement science evolves. Please provide recommendations to refine existing variables and their operational definitions included in the QMI, and any additional variables or domains CMS should consider for future inclusion in the QMI.** Consider providing an example of "other" in the operational definition of the measure submission method. **Q2. For the Reliability and Validity variables, please recommend additional tests or thresholds the QMI should consider acceptable.** None recommended. **Q3. Please provide feedback on the proposed scoring methodology, including the appropriateness of equal weighting of domains in calculating the overall QMI score.** None. **Q4. Since CMS has established health equity as a strategic priority, please comment on how the QMI might further incorporate health equity or disparities as a scoring variable.** Incomplete data on race, ethnicity and linguistic minority groups is a longstanding issue. Significant obstacles exist due to the lack of data for BIPOC and LGBTQ+ individuals. Consideration should also be given to the Intersectional implications among these groups. **Q5. To reduce burden on measure developers, CMS aims to closely align the QMI with National Quality Forum (NQF) evaluation criteria. Please provide recommendations to further align the QMI with NQF evaluation criteria.** Please consider solo private practice providers/small practices when looking at efforts to reduce collection burden as they have much fewer resources than larger healthcare systems and would benefit from a reporting system that is simple and user-friendly. **Q6. Please provide general recommendations for future considerations for the QMI.**

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	None.
5	June 13, 2022	Gregg Miller, MD CMO Vituity and MACRA-MDP TEP	Medical group	**Q1. The QMI can be adapted over time as strategic priorities change and measurement science evolves. Please provide recommendations to refine existing variables and their operational definitions included in the QMI, and any additional variables or domains CMS should consider for future inclusion in the QMI.** In this era of healthcare burnout and a focus on the fourth area of the Quadruple Aim, I was glad to see "Provider Burden" listed in Table 2 as a scoring variable. Measures should be evaluated based on their documentation burden to individual healthcare workers. Process & efficiency measures that require significant documentation in the EHR by a healthcare worker (eg SEP-1) should not score as highly as those that require minimal documentation. As a practicing physician, time spent documenting to meet a measure or to demonstrate why a particular case should be excluded from a measure takes away time I could be spending at the bedside. Even if a measure is able to demonstrate it uses the "least burdensome method available" it might still be too burdensome to approve. Perhaps this should be changed to, "least burdensome method is acceptably streamlined." **Q2. For the Reliability and Validity variables, please recommend additional tests or thresholds the QMI should consider acceptable.** No further recommendations. **Q3. Please provide feedback on the proposed scoring methodology, including the appropriateness of equal weighting of domains in calculating the overall QMI score.** Consider weighting "High Priority" more than other categories. For "Measure Performance," consider changing the scale for >5%/<5% to 10% or 20%. Many areas that could be improved in the hospital have significant opportunities, much greater than just 5%. **Q4. Since CMS has established health equity as a strategic priority, please comment on how the QMI might further incorporate health equity or disparities as a scoring variable.** Measures could be assessed as to how well they could be stratified by race/ethnicity, gender, etc to demonstrate whether there is care variation based on those factors.

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	**Q5. To reduce burden on measure developers, CMS aims to closely align the QMI with National Quality Forum (NQF) evaluation criteria. Please provide recommendations to further align the QMI with NQF evaluation criteria.** n/a **Q6. Please provide general recommendations for future considerations for the QMI.** While not as relevant to this iteration of the QMI, thought should be given to HCW-reported outcomes (not just patient-reported outcomes). Measures dealing with staffing levels, safety/workplace violence, appropriate clinical resources, engagement etc will be important if we are to improve system-based care.
6	June 13, 2022	Matthew K. Pickering, PharmD Senior Director, Measurement Science and Application National Quality Forum	NQF is an independent, nonprofit, 501(c)(3) membership organization that brings together diverse organizations and individuals from across the country dedicated to improving health and healthcare through quality measurement.	**Q1. The QMI can be adapted over time as strategic priorities change and measurement science evolves. Please provide recommendations to refine existing variables and their operational definitions included in the QMI, and any additional variables or domains CMS should consider for future inclusion in the QMI.** The National Quality Forum (NQF) appreciates the opportunity to comment on the Quality Measure Index (QMI) tool. NQF is a not-for-profit, nonpartisan, membership-based organization that works to catalyze improvements in healthcare. To that end, NQF convenes healthcare stakeholders from the public and private sectors with the aim to foster quality improvement through advancing important, valid, reliable, important and feasible measures through the application of consensus-based standards (i.e., quality measures). Today there are more than 400 NQF-endorsed measures used by more than 20 federal public reporting and pay-for-performance programs as well as in private-sector and state programs. NQF understands that the intent of the QMI tool is "to enhance, not replace, existing endorsement and measure selection processes," but to support priority setting within the Centers for Medicare & Medicaid Services (CMS). NQF appreciates that the QMI tool variables, variable components, and thresholds align well with NQF's standardized measure evaluation criteria, including the current preliminary work from its Scientific Methods Panel (SMP) (i.e., reliability thresholds recommendations). This concordance will help prioritize measures that are important, valid, and more likely to obtain NQF endorsement. Furthermore, NQF agrees that the tool uses objective information to determine an overall QMI score and appreciates the recognition that the scores for these variables should be interpreted as initial indicators that do not replace the multi-stakeholder expert reviews and public comment periods that provide wholistic, qualitative review of measure properties.

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	Regarding whether refinements are needed to existing variables or whether additional variables or domains should be considered, we recommend five areas in which the QMI tool may be enhanced: 1. NQF recognizes that health equity is fundamental to all quality improvement efforts. NQF applies an equity lens to every aspect of its work, with the goal of empowering healthcare stakeholders to take meaningful and measurable action to achieve health equity. This includes addressing quality and measurement gaps in key national health priorities, including the endorsement of performance measures that can identify and have the potential to reduce health disparities. Addressing the wide spectrum of disparities should be considered as a key component for successful health outcomes across the nation. Therefore, additional variables or domains within the QMI tool may focus on health equity and disparities. NQF is pleased to see that this consideration is within the list of future enhancements to the tool (page 17 of QMI report). In addition, advancing health equity has become a national priority and is highlighted in CMS' Quality Measurement Action Plan. If the QMI tool were to develop a new domain of health equity, it may be worth considering assigning a higher weight to these national priority domains, like health equity. 2. With respect to the scientific acceptability of measures (i.e., reliability and validity), NQF appreciates the attention towards the recommendations and ongoing work from NQF's SMP, specifically the "Draft Acceptability Reliability Thresholds Version 3.2". This draft framework of approaches to and acceptable thresholds for reliability is intended to aid measure stewards and developers in conducting reliability testing, preparing measure submissions, and assisting NQF committees with evaluating person-/encounter-level and accountable entity-level testing. However, this framework has not been finalized, nor incorporated into NQF's measure evaluation criteria. Therefore, NQF appreciates and further recommends the continued acknowledgement that this work is ongoing and that these thresholds are not established NQF criteria at this time. 3. For the risk adjustment variable, NQF agrees with the scoring approach, which is determining whether the measure is risk adjustment or risk stratified (e.g., outcome, resource use). However, on page 13, the scoring logic for the QMI tool appears to only assess whether risk adjustment or stratification was conducted but not the adequacy of adjustment (with rationale). NQF is currently developing technical guidance which will further inform how risk adjustment should be conducted with specific attention to social and functional risk, and when stratification of risk-adjusted measures is required. The standards and recommendations within this guidance will lead to future enhancements of NQF's measure evaluation criteria. We encourage CMS and HSAG to consider this work in the ongoing enhancements to the QMI tool.

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	4. For the feasibility and usability domain, NQF appreciates the attention towards the potential burden to providers. NQF agrees that this type of burden can manifest from the type of data being collected and reported, where digital data sources are more feasible than manual abstraction. However, the concept of provider burden may also stem from the complexity of the measure and whether there is a lack of harmonization of key measure elements (e.g., numerator, denominator, exclusions, risk adjustment models) to other similar measures currently used in federal programs. Additionally, the variable components for burden may also be expanded to consider any unintended consequences identified from the development and/or use of the measure. This is accordance to NQF's measure evaluation criteria, which evaluates whether the benefits of a performance measure in facilitating progress toward achieving high quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists). This evidence may be obtained via literature, feedback to the measure developer and/or measure implementer. Therefore, NQF encourages that future iterations of the QMI tool attempt to capture these additional consideration of burden. 5. Lastly, in the introduction of the QMI methodology report, there is reference to the Government Accountability Office (GAO) study that recommended "CMS develop and implement new procedures to systematically assess quality measures being considered for development, use, or removal from CMS programs to determine the impact on achieving CMS' strategic objectives." Further, the methodology report states that "the QMI tool attempts to address the GAO recommendations." With respect to the use or removal of measures from CMS programs, it is unclear if the QMI tool considered the standards and processes of NQF's Measures Applications Partnership (MAP). The MAP was created by section 3014 of the Patient Protection and Affordable Care Act (ACA) to provide recommendations to the Department of Health and Human Services (HHS) on the selection of performance measures for Medicare public reporting and performance-based payment programs. To conduct its work, the MAP uses its standardized Measure Selection Criteria (MSC) to guide its review of measures under consideration (MUCs) for inclusion into CMS programs. The MSC are intended to assist the MAP with identifying characteristics that are associated with ideal measure sets used for public reporting and payment programs. Similar to the QMI, the MSC are not absolute rules; rather, they are meant to provide general guidance on measure selection decisions and to supplement information provided on program-specific statutory and regulatory requirements. The central focus is on the selection of high-quality measures that address key national healthcare priorities. Preferences for measure selection include evaluating the relative strengths and weaknesses of a program measure

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	set and how the addition of an individual measure would contribute to the set. To determine whether a measure should be considered for a specified program, the MAP evaluates the MUCs against the MSC. Lastly, for the 2021-2022 MAP cycle, NQF has collaborated with CMS and the MAP to define a pilot process in which the MAP will recommend the removal of measures from CMS programs, referred to as the Measure Set Review (MSR). The MSR process also considers a series of standardized criteria that the MAP uses for its review and removal recommendations of measures currently within CMS programs. Due to the stated intent of the QMI tool, future enhancements may consider aligning with current MAP MSC and MSR criteria and processes. **Q2. For the Reliability and Validity variables, please recommend additional tests or thresholds the QMI should consider acceptable.** Please refer to the responses under prompt #1. However, measures at concept stage or still in development may not have testing results available. Therefore, these variables may not apply. **Q3. Please provide feedback on the proposed scoring methodology, including the appropriateness of equal weighting of domains in calculating the overall QMI score.** Please refer to recommendation #1 under prompt #1. Additionally, NQF defers to CMS policy preferences on the weighting and notes it may vary across measure types and programs. We recommend running CMS high and low priority measures through the QMI tool and confirming that they show up as high priority in the algorithm to test the face validity of the results using equal weighing. **Q4. Since CMS has established health equity as a strategic priority, please comment on how the QMI might further incorporate health equity or disparities as a scoring variable.** As stated previously, NQF applies a health equity lens to its work, which includes the endorsement of performance measures that can identify and have the potential to reduce health disparities. Therefore, additional variables or domains within the QMI tool may focus on health equity and disparities. This may include assessing whether the measure identifies gaps in quality of care between certain disadvantaged populations (i.e., stratifies performance scores by subgroups). NQF has previously developed criteria to identify disparities-sensitive measures, which are those measures that serve to detect not only differences in quality across institutions or in relation to certain benchmarks, but also differences in quality among populations or social groupings (race, ethnicity, language, etc.). Building from NQF's work in this area, the QMI tool may be enhanced to further prioritize individual performance measures that can serve to identify disparities in care. The concept of disparities-sensitive measures is also being considered by NQF's Core Quality Measure

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	Collaborative (CQMC). The CQMC is a partnership between AHIP and the CMS with involvement from NQF. More than 75 stakeholders comprise CQMC, including health insurers, medical associations, consumer groups, and others. These organizations are convened to identify core sets of quality measures and to identify quality measure gaps, including identifying current CQMC measures that are disparities-sensitive. This work is in progress and can be shared once completed. **Q5. To reduce burden on measure developers, CMS aims to closely align the QMI with National Quality Forum (NQF) evaluation criteria. Please provide recommendations to further align the QMI with NQF evaluation criteria.** Please refer to the responses under prompt #1 **Q6. Please provide general recommendations for future considerations for the QMI.** Please refer to the responses under prompt #1
7	June 15, 2022	Nancy Tuders, RN, IP-BC, GERO-BC, CDONA, FADONA, Master Trainer Assistant Director of Education NADONA National Association of Directors of Nursing Administration - - Springdale, OH	Resource and Education for LTC Leaders	**Q1. The QMI can be adapted over time as strategic priorities change and measurement science evolves. Please provide recommendations to refine existing variables and their operational definitions included in the QMI, and any additional variables or domains CMS should consider for future inclusion in the QMI.** We are dealing with a variety of QM's. Some for 5 Star, some for VBP, some for QRP, some in the Casper. DON's are leaving the industry due to lack of staff and the need to work as a floor nurse. They CANNOT KEEP TRACK of all the different QM's and which ones belong to what program. It is RIDICULOUS. We need ONE SET OF QM'S ACROSS THE BOARD. Those writing these measures need a reality check as to what is going on in facilities these days with simply trying to provide care. ONE SET would allow a stronger focus and I am certain, as a former DON, would result in improved measures. **Q2. For the Reliability and Validity variables, please recommend additional tests or thresholds the QMI should consider acceptable.** Should consider variences for the type of residents served. For example...a facility has an entire TBI Unit and they score high on psychotropic meds. These are often younger adults, and a period during their diagnosis benefit from psychotropics and often prevent further injury, just like a "net bed". They should not be lumped in with others receiving antipsychotics. **Q3. Please provide feedback on the proposed scoring methodology, including the appropriateness of equal weighting of domains in calculating the overall QMI score.**

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	see above **Q4. Since CMS has established health equity as a strategic priority, please comment on how the QMI might further incorporate health equity or disparities as a scoring variable.** see above **Q5. To reduce burden on measure developers, CMS aims to closely align the QMI with National Quality Forum (NQF) evaluation criteria. Please provide recommendations to further align the QMI with NQF evaluation criteria.** see above **Q6. Please provide general recommendations for future considerations for the QMI.** see above
8	June 20, 2022	Anders Chen, MD MHS Chair, SGIM Health Policy Clinical Practice Committee Society of General Internal Medicine	Physician Professional Society	**Q1. The QMI can be adapted over time as strategic priorities change and measurement science evolves. Please provide recommendations to refine existing variables and their operational definitions included in the QMI, and any additional variables or domains CMS should consider for future inclusion in the QMI.** Commenter skipped this question. **Q2. For the Reliability and Validity variables, please recommend additional tests or thresholds the QMI should consider acceptable.** Commenter skipped this question. **Q3. Please provide feedback on the proposed scoring methodology, including the appropriateness of equal weighting of domains in calculating the overall QMI score.** Commenter skipped this question. **Q4. Since CMS has established health equity as a strategic priority, please comment on how the QMI might further incorporate health equity or disparities as a scoring variable.** The Society of General Internal Medicine (SGIM) strongly believes that equity should be taken into account in measure selection. SGIM has previously advocated for creation of health equity specific quality measures – for example, providing language concordant services; screening for unmet health-related social needs; acting on these screens (placing referrals, completion of referrals and receipt of services), completion of community needs health assessments and action on these assessments. A scoring variable to assess a measures' potential to improve equity would be beneficial as

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	future equity measures are created. This is to some degree addressed in the High Priority scoring variable, but SGIM believes equity should be a separate scoring variable. **Q5. To reduce burden on measure developers, CMS aims to closely align the QMI with National Quality Forum (NQF) evaluation criteria. Please provide recommendations to further align the QMI with NQF evaluation criteria.** Commenter skipped this question. **Q6. Please provide general recommendations for future considerations for the QMI.** Commenter skipped this question.
9	June 20, 2022	Amanda Holt, MPH Senior Strategist, Healthcare Quality American Academy of Family Physicians (AAFP) and American Board of Family Medicine (ABFM)	Specialty Society & Specialty Licensing Board	**Q1. The QMI can be adapted over time as strategic priorities change and measurement science evolves. Please provide recommendations to refine existing variables and their operational definitions included in the QMI, and any additional variables or domains CMS should consider for future inclusion in the QMI.** The American Academy of Family Physicians (AAFP) and the American Board of Family Medicine (ABFM) strongly support including provider burden as a variable. Primary care physicians continue to cite measure reporting as one of the key drivers of administrative burden and burnout. Reducing burden by using data collected as a by-product of patient care, including the use of claims data when possible, should be central to CMS' approach for considering whether to include measures in its programs. The AAFP and ABFM appreciate CMS including risk adjustment as a variable. Risk adjustment and stratification are key to meeting our common goals of improving quality and mitigating health disparities. However, CMS notes the current QMI does not consider the approach used for risk adjustment. Thus, we encourage CMS to prioritize this in future iterations of the QMI. We also strongly urge CMS to incorporate alignment and attribution as variables in the future. Measure alignment across payers and quality programs is a key priority of the AAFP and numerous other stakeholders. Stronger preference should be given to measures that are used across more payers. That said, it's important to consider the level that each measure is specified and validated to measure. (i.e., We would not support the inclusion of a measure in the QPP that was created and specified at the health plan level.) In other words, just because a measure scores highly on a future alignment variable, does not mean it should be included in certain federal programs.

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	We also strongly support including evidence base as a variable. Incentives should be aligned for physicians and other clinicians to provide care in line with the latest clinical guidance, which is developed through robust examination of existing scientific evidence. CMS should not include measures in its programs that are inconsistent with USPSTF and national clinical guidelines developed by AAFP and our peer organizations. In future versions of the QMI, we encourage CMS to include a variable and/or variables that relate to value, recognizing that value is multifaceted: • Value to clinicians (they find the measure meaningful and worth their time and investment). • Value to patients (patients find the information meaningful and will assist them with evaluating groups of physicians)' • Value to purchasers (demonstrates their purchase contributes to the health and wellbeing of their employees); and • Value to payers (demonstrates meaningful ROI). **Q2. For the Reliability and Validity variables, please recommend additional tests or thresholds the QMI should consider acceptable.** Reliability and validity testing should be performed at the level of intended measurement. (i.e., tested at the clinician level for clinician reporting programs, group/practice level for group/practice reporting, etc.) As mentioned above, it's important to consider the level that each measure is specified and validated to measure. **Q3. Please provide feedback on the proposed scoring methodology, including the appropriateness of equal weighting of domains in calculating the overall QMI score.** The American Academy of Family Physicians (AAFP) and the American Board of Family Medicine (ABFM) disagree with the proposed equal weighting across domains. Measures that score poorly on the feasibility & usability domain or the importance domain (or both) should not be added to CMS programs regardless of how they score across other domains. We believe this approach would be consistent with CMS meaningful measures initiative and would facilitate evidence-based care across CMS programs. We would prefer to see feasibility/usability and importance domains carry more weight than the scientific acceptability domain. For example, if a measure reaches a certain threshold on these two domains, then the scientific acceptability domain becomes relevant. If the scientific acceptability of a measure is high but the collection of the

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	data to calculate the measure is not feasible or highly burdensome or if the measure itself is not important, i.e., meaningful, then the measure should not be included. We also reiterate our recommendation for cross-payer alignment to be included in the scoring methodology. The AAFP and ABFM request that CMS outline how these scores will be used to determine measure inclusion in various programs. Establishing appropriate and reliable scoring thresholds and seeking public comment on these thresholds is an important step for ensuring the QMI fulfills its intended purpose. **Q4. Since CMS has established health equity as a strategic priority, please comment on how the QMI might further incorporate health equity or disparities as a scoring variable.** There could be different approaches to further incorporate health equity or disparities as scoring variables. One approach is demonstrating that either the measure concept or how the measure is constructed, leads to improved equity in care or measurement. Regardless of the approach, the overarching goal should be to drive quality improvement of care for historically marginalized and medically underserved populations. At a more granular level, it's important to have quality measures that help address social determinants of health and social needs (food insecurity, housing, transportation, access to care, etc.) These types of measures should include automatic, seamless data collection of demographic and socioeconomic information like race, ethnicity, language, gender, dual status, etc. CMS should explore the use of this data so that clinicians caring for disadvantaged populations are assessed for their ability to raise quality, rather than the expectation that they can achieve equity. **Q5. To reduce burden on measure developers, CMS aims to closely align the QMI with National Quality Forum (NQF) evaluation criteria. Please provide recommendations to further align the QMI with NQF evaluation criteria.** The AAFP and ABFM appreciate the goal of aligning with NQF evaluation criteria, but it's important to recognize that such criteria are often subjective and lack a consistent process. Attention should be given to make evaluation criteria concrete, transparent, and consistent. If CMS aims to gather evaluation criteria from NQF submission forms, the data may be outdated and inaccurate, depending on how long the measure has been around. As

Comment Number	Date Posted/ Received	Name, Credentials, and Organization of Commenter	Type of Organization	Text of Comments
(cont.)	(cont.)	(cont.)	(cont.)	such, we recommend CMS work with NQF to develop a process that would allow CMS to update the QMI at the same time as NQF is modifying evaluation criteria. We also strongly encourage CMS to use the least-burdensome data sources for quality measurement. In many instances, claims data is less burdensome than EHR data. **Q6. Please provide general recommendations for future considerations for the QMI.** The current 1.0 version of the QMI is a good start. The AAFP and ABFM support the overall intent of this new tool. However, the QMI needs further revision to meet its intended purpose. As listed on page 16 of the QMI Methodology Report, there are several limitations of this initial version that need to be acknowledged and rectified in future versions. It's important to keep in mind that the QMI is intended to serve as a supplement — not a replacement — for all other stakeholder review and feedback opportunities. For instance, the Measure Applications Partnership (MAP), convened by NQF, recently solicited comments on a list of 32 proposed measures for removal from federal programs. We recommend that CMS continue to value recommendations provided by the MAP and other stakeholders. We also question the idea CMS poses about including avoided costs in future iterations of the index. It is imperative that CMS programs consider cost only when looking at quality simultaneously. Measures should therefore be assessed on their ability to improve quality of care. Reducing cost should never take precedence over whether a measure is meaningful, feasible, evidence based. Additionally, we recommend the QMI consider whether or not measures include patient input. The AAFP and ABFM share CMS' belief that incorporating the patient perspective into quality improvement is vital. The QMI should promote the inclusion of the patient experience in quality measurement. Finally, the AAFP and ABFM request that CMS provide more information about how it expects the QMI to be used, including details about who should use it, when they should use it, and what types of decisions should be made based on QMI data.

Note: The text of comments was entered verbatim without edits for spelling, punctuation, grammar, or any other reason.