



Quality Measure Index (QMI) Methodology Report

Release, Version 1.0

April 29, 2022

Prepared for the Centers for Medicare & Medicaid Services
by Health Services Advisory Group, Inc.



Table of Contents

Introduction.....	2
QMI Development and Testing	3
External Stakeholder Input.....	3
Literature Review	3
QMI Testing	3
QMI Variables, Scoring, and Interpretation	4
Variables.....	4
Scoring	6
Variable-Level Scoring.....	6
Domain-Level and Overall Scoring.....	14
Interpretation	15
QMI Tool and Abstraction Guide.....	15
Limitations	16
Next Steps	17
Conclusion	17
References.....	18
Appendix A – Variable Scoring Logic	19
Importance Domain.....	19
Scientific Acceptability Domain	22
Feasibility and Usability Domain.....	29
Appendix B – Measure Scoring Example.....	31

Quality Measure Index (QMI) Methodology Report

Introduction

Currently, information to assess quality measures for use in the Centers for Medicare & Medicaid Services (CMS) quality programs is heterogeneous, imprecise, and lacks standardization. These factors have inhibited the fair comparison of the relative value of quality measures in achieving CMS strategic objectives. CMS contracted with HSAG (“the team”) to develop the Quality Measure Index (QMI) to address these limitations. The QMI is a tool intended to support the assessment and selection of quality measures that provide meaningful quality performance information, align with the Healthcare Quality Priorities of the Meaningful Measures Framework, and fulfill requirements of the Medicare Access and Children’s Health Insurance Program (CHIP) Reauthorization Act of 2015 (MACRA) by supporting CMS’ evaluation and selection of measures that reflect current priorities.

In September 2019, the Government Accountability Office (GAO) published a study that found CMS used various approaches to identify measures to be developed for, used in, or removed from quality programs.¹ The report noted that CMS lacked a systematic method to assess measures to ensure they meet CMS strategic objectives. Therefore, the GAO recommended that 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. The QMI aims to address the GAO recommendations.

The QMI is envisioned as a transparent and reliable scoring instrument based on standardized definitions of quantifiable measure characteristics. Capable of producing repeatable results yet adaptable to evolving priorities, the index thus provides capabilities that are unique among current assessment tools used in decision-making. The goal of the QMI is to systematically and transparently display the strengths and limitations of each quality measure to facilitate comparisons among measures and aid CMS in selecting high-value measures that relate to strategic agency priorities (e.g., reduction of burden, improving health outcomes, and health equity) for program implementation in quality reporting programs. The tool also can be used at various stages of measure development to identify potential limitations of a given measure and inform decisions about when to discontinue development, when the measure requires re-specification, and when the measure is ready for consensus-based entity (CBE) endorsement consideration (e.g., through the National Quality Forum (NQF)).

The QMI is intended to enhance, not replace, existing endorsement and measure selection processes. The QMI fills a critical gap, providing CMS a method to assess measures based on objective criteria, and can complement qualitative expert reviews of measure information through public comment, CBE endorsement, and pre-rulemaking.

The purpose of this methodology report is to provide an overview of the QMI development approach and describe the QMI methodology used to assess and score facility-level and clinician-level measures, which includes the QMI variables and domains and their associated scoring.

QMI Development and Testing

The QMI is composed of key measure characteristics that reflect the quality of a measure. These characteristics are operationalized as variables, each with its own operational definition and scoring approach. The team conducted a series of activities to identify these variables, determine feasible operational definitions reflective of each variable’s intent, and define a scoring approach at the variable and domain level. External stakeholder input, literature reviews, and alpha and beta testing completed to date are described in subsequent sections.

External Stakeholder Input

External stakeholder input has played a vital role in development of the QMI. Throughout the development of the QMI, the team has presented developmental milestones and solicited feedback from external stakeholders (e.g., Technical Expert Panels [TEPs]) and CMS leadership and staff.

Stakeholders provided input on QMI variable assessment and feasibility, scoring and weighting, and ongoing refinements to the index. In 2022, the team plans to continue discussions with the TEP and external stakeholders regarding further refinement of the QMI. This methodology document will be used to solicit public comment in spring 2022.

Literature Review

The team comprehensively reviewed literature to support QMI development. The comprehensive literature reviews focused on the evaluation and critique of quality measures. Sources included, but were not limited to, foundational industry standards (e.g., *CMS Measure Management System (MMS) Blueprint* (the “Blueprint”)², National Quality Forum (NQF) Measure Evaluation Criteria³), priorities from MACRA and the CMS Measure Development Plan (MDP), industry reports related to quality measure development, and peer-reviewed literature.

Through literature reviews the team identified measure characteristics and variables across levels of analysis and phases of the Measure Lifecycle. The applicability and feasibility of QMI variables for measure assessment across CMS programs and settings in all phases of the Measure Lifecycle were also assessed. Additionally, the literature reviews found sustained applicability of the QMI variables across settings and programs and informed refinements to the variable operational definitions and scoring approaches.

Lastly, some measure characteristics identified through the review of literature or proposed by external stakeholders were removed from the QMI due to a lack of consistent definitions. For instance, variables such as *Alignment* and *Attribution* were removed because standardized and widely accepted operational definitions do not yet exist for these concepts. The removed QMI variables, however, represent important characteristics of quality measures and could be added to the index in the future.

QMI Testing

The team has conducted multiple rounds of testing using measures in use in CMS programs, newly developed measures, and measures still in development. Findings suggested the QMI variables can be

applied across settings and levels of analysis, and across the five phases of the Measure Lifecycle.¹ Testing confirmed the feasibility, reliability, and validity of the scoring variables. Throughout testing, quality measure documentation showed substantial heterogeneity.

To standardize measure information needed to calculate QMI scores, the team collected additional information through the 2021 and 2022 Pre-Rulemaking processes. Measure information submitted during the 2022 Pre-Rulemaking process will be reviewed to identify further revisions or refinements to improve standardization and requirements for measure information submission.

QMI Variables, Scoring, and Interpretation

The following section describes the QMI variables, domains, and scoring.

Variables

QMI variables are grouped as classification or scoring variables. Classification variables are used to group or stratify quality measures and are not incorporated into the QMI measure score. The eight QMI classification variables and their operational definitions are shown in Table 1 below.

Table 1. QMI Classification Variables and Operational Definitions

Variable Name	Operational Definition
Meaningful Measures Classification	Applicable Meaningful Measures 2.0: Person-Centered Care, Equity, Safety, Affordability and Efficiency, Chronic Conditions, Wellness and Prevention, Seamless Care Coordination, Behavioral Health*
Measure Type	Applicable measure type: Outcome, Patient-Reported Outcome-Based Performance Measure (PRO-PM), Intermediate outcome, Process, Efficiency, Structure, Other.
Composite Measure	The measure is classified as a composite or a component of a composite.
Measure Submission Method	The method in which the measure is submitted to CMS: Claims, Web-based tool, Electronic clinical quality measures (eCQM), Clinical quality measure (CQM), Qualified Clinical Data Registry (QCDR), Other.
NQF Endorsement Status	The measure's NQF endorsement status: Endorsed, Endorsed-Reserve, Failed Endorsement, Endorsement Removed, eCQM Approved for Trial Use (Not Endorsed), Not Endorsed (No NQF form).
Development Phase	The measure's completed phase of development: conceptualization, specification, testing, implementation, or use, continuing evaluation, and maintenance**.
Digital Measure	The measure 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 [§] .
Core Quality Measures Collaborative (CQMC) Measure Sets	The measure is from one of 10 published CQMC Core Measure Sets: - Accountable Care Organizations/Patient Centered Medical Homes/Primary Care - Behavioral Health - Cardiology - Gastroenterology - HIV & Hepatitis C - Medical Oncology - Neurology - Obstetrics & Gynecology - Orthopedics - Pediatrics

*Due to availability of data, Meaningful Measures 1.0 was utilized during all rounds of testing to date. For the 2022 measure assessment, the classification variable has been updated to reflect Meaningful Measures 2.0. **Measure development phases as specified in the CMS MMS Blueprint. [§]Digital Measure definition to be updated to align with CMS draft definition published 4/11/22.

¹ The five phases of the Measure Lifecycle are conceptualization; specification; testing; implementation; and, use, continuing evaluation, and maintenance, as defined by the Blueprint, V17.0.²

Scoring variables reflect data integral to the quantitative assessment of quality measures and are used to calculate the QMI measure score. Each scoring variable is intended to assess a single characteristic of the measure and is grouped into one of three domains: Importance, Scientific Acceptability, and Feasibility and Usability. Variable components are distinct elements that are integral to operationalizing a specific QMI scoring variable. Variable components are only pertinent to the *Reliability* and *Validity* variables where each component represents a specific level of testing. The eight QMI scoring variables, organized by domains, and their operational definitions are included in Table 2.

Table 2. QMI Scoring Variables, Domain Names, and Operational Definitions

Variable Name	Variable Component	Operational Definition
Importance Domain		
<i>Evidence-Based</i>	Not applicable	Outcome measures provided at least one citation for evidence and all other types of measures are supported by a clinical guideline, systematic review of the literature, and/or other evidence from the literature or empiric analyses.
<i>High Priority</i>	Not applicable	The measure's number of high priorities as defined by three CMS strategic measurement priorities: outcome, intermediate outcome, or PRO-PM; digital; and equity.*
<i>Measure Performance</i>	Not applicable	The relative difference between the mean of accountable entity (e.g., facility, clinician) performance scores and a performance benchmark.
Scientific Acceptability		
<i>Reliability</i>	Measure Score	Signal-to-noise or random split-half correlation results using measure scores for the same level of analysis as the intended measure use
	Data Element	Interrater reliability results for each of the data elements required to score the measure.
	Survey-Level Testing	Surveys and patient-reported outcome measures (PROM) included in the performance measure have undergone psychometric testing
<i>Validity</i>	Empiric Measure Score	Empirical correlation using measure scores for the same types of measured entities as the intended measure use.
	Data Element	Agreement with an authoritative source for each of the data elements required to score the measure.
	Face Validity	The majority of experts consulted agree the measure has face validity.
	Survey-Level Testing	Surveys and patient-reported outcome measures (PROM) included in the performance measure have undergone psychometric testing.
<i>Risk Adjustment</i>	Not applicable	Outcome, PRO-PM, intermediate outcome, and resource use measures are risk adjusted
Feasibility & Usability Domain		
<i>Feasibility</i>	Not applicable	To what extent are the specified data elements available electronically in defined fields (i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields).
<i>Provider Burden</i>	Not applicable	The least burdensome method available to providers to calculate measure scores each reporting period

*High Priority definition subject to change as CMS priorities evolve.

Scoring

This section describes the approach to scoring each variable and the domains and calculating the overall QMI score for each measure. For details on individual variable scoring logic, see *Variable Scoring Logic* (Appendix A). Scoring of a sample measure is shown in *Measure Scoring Example* (Appendix B).

Variable-Level Scoring

Each QMI variable was scored on a scale from 0.00 to 1.00, with higher scores indicating better quality for that measure characteristic. Color-coded variable-level scores help provide a visual appreciation of a measure’s strengths and limitations. The scores were qualitatively distinguished from each other through the color scheme shown in Table 3.

Table 3. Scoring Interpretation

Scoring Categories	Interpretation*	Scoring Value
Green	Preferred	1.00
Yellow	Acceptable	0.75
Red	Not preferred	0.25
Grey	Unable to determine due to missing information	0.00

* The interpretations are subject to change.

After review with the TEP, scoring values of 0.75 and 0.25 were chosen for the “Acceptable” and “Not Preferred” scoring categories, respectively. These values are considered to have face validity in terms of their contribution to the overall QMI scores. The values are designed to give similar and heavier weight to “Acceptable” and “Preferred” responses relative to “Not Preferred” and “Missing” data. Additionally, this approach promotes variation in measures’ overall QMI scores. Measures with missing information for a given variable score 0.00 for that variable to reflect that the variable cannot be evaluated and to encourage submission of complete measure information.

Table 4 provides the scoring approach for each variable and the considerations or limitations when scoring. For two scoring variables, *Reliability* and *Validity*, multiple components are collected and assessed. For these two variables, a hierarchy determines the component used for scoring. This determination is based on the type of available testing results that best reflect the overall performance of a measure. The rationale for this approach is that only one level of testing is required by NQF and CMS.^{2,3}

Table 4. Variables with Scoring Approach

Variable Name	Scoring Approach Considerations and Limitations								
Importance Domain									
Evidence-Based	<p>Variable Scoring Approach: The QMI score for this variable is based on the strongest evidence provided by the developer. When more than one citation for evidence is provided to support a measure, the QMI score gives priority to graded US guidelines.</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="background-color: #d9ead3;">•Outcome measures with at least one citation •Other measures with: •Strong or moderate guideline recommendation OR •U.S. Preventive Services Task Force (USPSTF) grade A, B, or D</td> <td style="text-align: right; vertical-align: bottom;">1.00</td> </tr> <tr> <td style="background-color: #fff2cc;">•Other measures with: •Guideline based on expert opinion OR •Systematic review without guideline recommendation OR •Ungraded guideline</td> <td style="text-align: right; vertical-align: bottom;">0.75</td> </tr> <tr> <td style="background-color: #f4cccc;">•Outcome measures without at least one citation •Other measures with: •Conditional or weak guideline recommendation OR •USPSTF grade C or I, OR •Cited literature without systematic review OR •Empiric data or other evidence</td> <td style="text-align: right; vertical-align: bottom;">0.25</td> </tr> <tr> <td style="background-color: #e6e6e6;">•No evidence provided</td> <td style="text-align: right; vertical-align: bottom;">0.00</td> </tr> </table> <p>Limitations: Since the variable score is based on the single best type of cited evidence provided, there is a possibility the variable score could overestimate the strength of evidence if the best evidence was not directly related to the measure concept and weaker evidence was more directly related to the measure concept. In rare instances, the best evidence for a measure is considered qualitatively weak but does provide direct support to the measure concept (i.e., inappropriate use measures citing a weak or conditional guideline recommendation as justification to discourage a process). Therefore, the score for this variable should be interpreted as an initial indicator that does not replace the need for secondary review of a measure's evidence base.</p>	•Outcome measures with at least one citation •Other measures with: •Strong or moderate guideline recommendation OR •U.S. Preventive Services Task Force (USPSTF) grade A, B, or D	1.00	•Other measures with: •Guideline based on expert opinion OR •Systematic review without guideline recommendation OR •Ungraded guideline	0.75	•Outcome measures without at least one citation •Other measures with: •Conditional or weak guideline recommendation OR •USPSTF grade C or I, OR •Cited literature without systematic review OR •Empiric data or other evidence	0.25	•No evidence provided	0.00
•Outcome measures with at least one citation •Other measures with: •Strong or moderate guideline recommendation OR •U.S. Preventive Services Task Force (USPSTF) grade A, B, or D	1.00								
•Other measures with: •Guideline based on expert opinion OR •Systematic review without guideline recommendation OR •Ungraded guideline	0.75								
•Outcome measures without at least one citation •Other measures with: •Conditional or weak guideline recommendation OR •USPSTF grade C or I, OR •Cited literature without systematic review OR •Empiric data or other evidence	0.25								
•No evidence provided	0.00								
High Priority	<p>Variable Scoring Approach: The QMI tabulates the number of CMS priorities for measurement a quality measure addresses. At this time a measure can address a maximum of three priorities: by being an outcome measure, intermediate outcome measure or PRO-PM; a digital measure based on the 2021 CMS definitionⁱⁱ and/or a measure that addresses health equity based on Meaningful Measures 1.0 Healthcare Quality Area “Equity of Care.”</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="background-color: #d9ead3;">•2+ priorities</td> <td style="text-align: right; vertical-align: bottom;">1.00</td> </tr> <tr> <td style="background-color: #fff2cc;">•1 priority</td> <td style="text-align: right; vertical-align: bottom;">0.75</td> </tr> <tr> <td style="background-color: #f4cccc;">•No priority</td> <td style="text-align: right; vertical-align: bottom;">0.25</td> </tr> <tr> <td style="background-color: #e6e6e6;">•Not applicable</td> <td style="text-align: right; vertical-align: bottom;">N/A</td> </tr> </table>	•2+ priorities	1.00	•1 priority	0.75	•No priority	0.25	•Not applicable	N/A
•2+ priorities	1.00								
•1 priority	0.75								
•No priority	0.25								
•Not applicable	N/A								

ⁱⁱ 85 FR 84849

Variable Name	Scoring Approach Considerations and Limitations								
	<p>Limitations: Given that CMS’ definition of digital measures continues to evolve, the QMI uses the most generous definition based on the data sources used.</p> <p>The QMI, to date, evaluated measures based on Meaningful Measures 1.0 which consists of 6 different priorities and 19 measurement areas, of which Equity of Care is a measurement area captured under the "Work with Communities to Promote Best Practices of Healthy Living" Priority. With the update of the variable to reflect Meaningful Measures 2.0, which consists of 8 domains, including one domain that captures measures designed to promote equity in health care, measures may be classified differently. In addition, measure information sources from which data for the QMI are abstracted present a single quality priority per measure; therefore, measures that address health equity as a secondary priority will not be captured.</p>								
Measure Performance	<p>Variable Scoring Approach: Measures with larger relative differences receive higher scores on this variable because larger differences are an indication of greater opportunity for improvement. If a distribution of accountable entity (e.g., facility, clinician) scores is available, the benchmark for "higher score is better" measures is the maximum score in the distribution and the benchmark for "lower score is better" measures is the minimum score in the distribution. If a distribution of accountable entity scores is not available for proportion measures, the QMI assigns the theoretical maximum of 100 to "higher score is better" measures and a theoretical minimum of 0 to "lower score is better" measures. Non-proportion measures that did not provide a distribution of accountable entity scores receive the lowest score for this variable because a theoretical minimum or maximum cannot be applied consistently across those measures and the relative room for improvement is unknown. Similarly, 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. A threshold of 5% relative difference is used to distinguish a QMI score of 1.00 (preferred) from 0.25 (not preferred).</p> <table border="1" data-bbox="381 1276 1291 1444"> <tbody> <tr> <td>•\geq 5% relative room for improvement</td> <td>1.00</td> </tr> <tr> <td>•Not applicable</td> <td>N/A</td> </tr> <tr> <td>•$<$ 5% relative room for improvement</td> <td>0.25</td> </tr> <tr> <td>•Unable to determine based on missing mean/median and/or benchmark</td> <td>0.00</td> </tr> </tbody> </table> <p>Limitations: By using the extremes of the accountable entity score distribution (i.e., minimum or maximum depending on whether a lower or higher score was better), there is potential for outlier entities to be selected as the benchmark and for room for improvement to be overestimated. This approach was selected due to limited and inconsistent data availability at other percentiles in the distributions and to avoid the risk of underestimating room for improvement. Additionally, by setting the benchmark at 100% or 0% for proportion measures that did not provide a distribution of accountable entity scores, this variable may overestimate the room for improvement for some proportion measures. However, we anticipate the impact of this imputation would be minimal because testing of proportion measures for which a distribution was available showed that use of 100% or 0% as a theoretical benchmark instead of the minimum or maximum in the distribution did not change the QMI <i>Measure Performance</i> variable score for those measures.</p>	• \geq 5% relative room for improvement	1.00	•Not applicable	N/A	• $<$ 5% relative room for improvement	0.25	•Unable to determine based on missing mean/median and/or benchmark	0.00
• \geq 5% relative room for improvement	1.00								
•Not applicable	N/A								
• $<$ 5% relative room for improvement	0.25								
•Unable to determine based on missing mean/median and/or benchmark	0.00								

Variable Name	Scoring Approach Considerations and Limitations								
Scientific Acceptability Domain									
Reliability: Measure Score Component	<p>Variable Scoring Approach: When provided, measure score reliability testing determines the score for the <i>Reliability</i> variable in the Scientific Acceptability domain.</p> <p>Measure score-level reliability testing results provide information on whether the performance measure scores for measured entities produce the same results a high proportion of the time when assessed in the same population and the same time period. The minimum threshold of 0.5 is aligned with the NQF Scientific Methods Panel (SMP) recommended threshold proposed as of July 2021.⁴</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="background-color: #d9ead3;">• ≥ 0.80 • Large effect size for eta-squared</td> <td style="text-align: center;">1.00</td> </tr> <tr> <td style="background-color: #fff2cc;">• < 0.80 and ≥ 0.5 • Moderate effect size for eta-squared analysis</td> <td style="text-align: center;">0.75</td> </tr> <tr> <td style="background-color: #f4cccc;">• < 0.50 • Small effect size for eta-squared</td> <td style="text-align: center;">0.25</td> </tr> <tr> <td style="background-color: #cccccc;">• Missing or different type of analysis provided</td> <td style="text-align: center;">0.00</td> </tr> </table> <p>Limitations: The heterogeneity in measure score reliability results provided in measure information sources used to calculate QMI scores resulted in the QMI allowing for eta-squared analysis for a small number of measures that provided that statistic. In the future, we anticipate that clearer guidance from NQF on expectations for these analyses and a new version of MERIT that is aligned with QMI variable requirements will encourage developers to use the accepted signal-to-noise or random split-half analyses to ensure further standardization in testing approaches and documentation practices.</p>	• ≥ 0.80 • Large effect size for eta-squared	1.00	• < 0.80 and ≥ 0.5 • Moderate effect size for eta-squared analysis	0.75	• < 0.50 • Small effect size for eta-squared	0.25	• Missing or different type of analysis provided	0.00
• ≥ 0.80 • Large effect size for eta-squared	1.00								
• < 0.80 and ≥ 0.5 • Moderate effect size for eta-squared analysis	0.75								
• < 0.50 • Small effect size for eta-squared	0.25								
• Missing or different type of analysis provided	0.00								
Reliability: Data Element Component	<p>Variable Scoring Approach: If measure score reliability is not provided, the Scientific Acceptability domain score for <i>Reliability</i> is based on the data element component.</p> <p>Given that data element reliability does not provide information on the measure scores that would be reported by CMS, data element reliability scores are capped at 0.75 (acceptable), which is aligned with NQF's approach of capping the reliability score at moderate if this is the only information provided. The thresholds for Kappa, Intraclass Correlation (ICC), and Pearson are aligned with the thresholds proposed by the NQF SMP as of July 2021. The QMI uses the result for the least reliable data element because this ensures that all data elements are in the acceptable range for reliability.</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="background-color: #d9ead3;">• Not applicable</td> <td style="text-align: center;">N/A</td> </tr> <tr> <td style="background-color: #fff2cc;">• Above or equal to the NQF thresholds for Kappa (≥ 0.4), ICC (≥ 0.5), or Pearson (≥ 0.6)</td> <td style="text-align: center;">0.75</td> </tr> <tr> <td style="background-color: #f4cccc;">• Percent agreement, PPV, or Sensitivity: $\geq 70\%$</td> <td style="text-align: center;">0.25</td> </tr> <tr> <td style="background-color: #cccccc;">• Missing or different type of analysis provided</td> <td style="text-align: center;">0.00</td> </tr> </table> <p>Limitations: The heterogeneity in data element level reliability results provided in the measure information sources used to calculate the QMI scores resulted in the QMI accepting a variety of statistic types for this variable. While Kappa is generally the preferred statistic for data element reliability, there are instances where it cannot be calculated (e.g., when there is 100%</p>	• Not applicable	N/A	• Above or equal to the NQF thresholds for Kappa (≥ 0.4), ICC (≥ 0.5), or Pearson (≥ 0.6)	0.75	• Percent agreement, PPV, or Sensitivity: $\geq 70\%$	0.25	• Missing or different type of analysis provided	0.00
• Not applicable	N/A								
• Above or equal to the NQF thresholds for Kappa (≥ 0.4), ICC (≥ 0.5), or Pearson (≥ 0.6)	0.75								
• Percent agreement, PPV, or Sensitivity: $\geq 70\%$	0.25								
• Missing or different type of analysis provided	0.00								

Variable Name	Scoring Approach Considerations and Limitations								
	<p>agreement). Additionally, NQF allows for data element validity to count for data element reliability. Therefore, the QMI allows for percent agreement, positive predictive value, and sensitivity for data element reliability to account for scenarios where Kappa could not be calculated, or where the developer only provided data element validity testing results.</p> <p>Additionally, the heterogeneity in responses to the data element-level testing questions was a factor in selecting the lowest data element result for the QMI score. Some developers did not provide individual data element-level results and only provided results for an overall numerator or broken down by some other strata (e.g., state). The QMI accepts overall results and selects the lowest value where available. In the future, we anticipate that clearer guidance on expectations for this level of testing will encourage developers to provide results for individual data elements to allow for better comparison of results across measures. In addition, a new version of MERIT that is aligned with QMI variables may also enhance documentation practices.</p>								
<p>Reliability:</p> <p>Survey-Level Testing Component</p>	<p>Variable Scoring Approach: The Scientific Acceptability domain score for <i>Reliability</i> is determined by the survey-level testing component only if a measure reported survey-level testing data and did not provide measure score reliability, data element reliability or data element validity testing results of the performance measure itself. The maximum survey-level testing score is capped at 0.75 (acceptable) to give lower priority to measures that only conducted survey-level testing and did not provide information on performance measure-level testing.</p> <table border="1" data-bbox="381 1108 1286 1245"> <tbody> <tr> <td>•Not applicable</td> <td>N/A</td> </tr> <tr> <td>•Psychometric testing on survey or PROM conducted</td> <td>0.75</td> </tr> <tr> <td>•Not applicable</td> <td>N/A</td> </tr> <tr> <td>•Psychometric testing on survey or PROM not conducted</td> <td>0.00</td> </tr> </tbody> </table> <p>Limitations: Current measure information documentation provided by NQF and CMS include survey-level testing as options in the reliability field rather than in discrete fields. The QMI gives some credit when this information is provided in lieu of measure-level testing because developers may not have been aware that they should also provide performance measure testing results. The QMI also assumes the developer provides the survey-level testing that was appropriate, and the testing was applicable to the survey as specified in the performance measure. If more than one survey or PROM is included in the measure specifications, the QMI does not confirm that all surveys or PROMs have undergone psychometric testing. Therefore, the score for this variable should be interpreted as an initial indicator that does not replace the need for secondary review of a measure's survey-level testing.</p>	•Not applicable	N/A	•Psychometric testing on survey or PROM conducted	0.75	•Not applicable	N/A	•Psychometric testing on survey or PROM not conducted	0.00
•Not applicable	N/A								
•Psychometric testing on survey or PROM conducted	0.75								
•Not applicable	N/A								
•Psychometric testing on survey or PROM not conducted	0.00								
<p>Validity:</p> <p>Empiric Measure Score Component</p>	<p>Variable Scoring Approach: When provided, empiric measure score validity testing determines the QMI score for the <i>Validity</i> variable in the Scientific Acceptability domain. This variable provides information on the extent to which the performance measure scores for measured entities quantify what they purport to measure.</p>								

Variable Name	Scoring Approach Considerations and Limitations										
	<table border="1" data-bbox="381 386 1291 520"> <tr> <td>•At least one correlation in hypothesized direction</td> <td>1.00</td> </tr> <tr> <td>•Not applicable</td> <td>0.75</td> </tr> <tr> <td>•No correlation in hypothesized direction</td> <td>0.25</td> </tr> <tr> <td>•Missing or different type of analysis provided</td> <td>0.00</td> </tr> </table> <p>Limitations: Given that correlations are typically low for these types of analyses because of the inability to account for other factors that could influence the results, the QMI assumes that any results that are correlated in the hypothesized direction are an indication of measure validity regardless of the magnitude of the correlation or if the developer provided other correlations that were not in the hypothesized direction. The score is based on the best correlation provided and assumes the developer conducted the testing appropriate for the measure. This may result in an overestimate of measure score validity. Therefore, the score for this variable should be interpreted as an initial indicator that does not replace the need for qualitative secondary expert review of a measure's empirical validity testing.</p>	•At least one correlation in hypothesized direction	1.00	•Not applicable	0.75	•No correlation in hypothesized direction	0.25	•Missing or different type of analysis provided	0.00		
•At least one correlation in hypothesized direction	1.00										
•Not applicable	0.75										
•No correlation in hypothesized direction	0.25										
•Missing or different type of analysis provided	0.00										
<p>Validity:</p> <p>Data Element Component</p>	<p>Variable Scoring Approach: If measure score validity is not provided, the Scientific Acceptability domain score for <i>Validity</i> is based on the data element validity component. Given that data element validity does not provide information on the measure scores that would be reported by CMS, the data element validity scores are capped at 0.75 (acceptable), which is aligned with NQF's approach of capping the validity score at moderate if this is the only information provided. The QMI uses the result for the lowest data element because this ensures that all data elements are in the acceptable range for validity.</p> <table border="1" data-bbox="381 1115 1291 1310"> <tr> <td>•Not applicable</td> <td>N/A</td> </tr> <tr> <td>•Above or equal to the NQF thresholds for Kappa (≥ 0.4), ICC (≥ 0.5), or Pearson (≥ 0.6)</td> <td>0.75</td> </tr> <tr> <td>•Percent agreement, PPV, or Sensitivity: $\geq 70\%$</td> <td></td> </tr> <tr> <td>•Below threshold for relevant statistic type</td> <td>0.25</td> </tr> <tr> <td>•Missing or different type of analysis provided</td> <td>0.00</td> </tr> </table> <p>Limitations: While Positive Predictive Value (PPV) or Sensitivity are generally the preferred statistic for data element validity, it can require significantly more resources to conduct those types of analyses due to the volume of data required and the need for gold standard comparison. Therefore, the QMI accepts additional statistics such as percent agreement and Kappa if they were provided in the data element validity sections of the forms.</p> <p>Additionally, the heterogeneity in responses to the data element level testing questions was a factor in selecting the lowest data element result for the QMI score. Some developers did not provide individual data element-level results and only provided results for an overall numerator or broken down by some other strata (e.g., state). The QMI accepts overall results and selects the lowest value where available. In the future, we anticipate that clearer guidance on expectations for this level of testing will encourage developers to provide results for individual data elements to allow for better comparison of results across measures. In addition, a new version of MERIT that is aligned with QMI variables may also enhance documentation standards.</p>	•Not applicable	N/A	•Above or equal to the NQF thresholds for Kappa (≥ 0.4), ICC (≥ 0.5), or Pearson (≥ 0.6)	0.75	•Percent agreement, PPV, or Sensitivity: $\geq 70\%$		•Below threshold for relevant statistic type	0.25	•Missing or different type of analysis provided	0.00
•Not applicable	N/A										
•Above or equal to the NQF thresholds for Kappa (≥ 0.4), ICC (≥ 0.5), or Pearson (≥ 0.6)	0.75										
•Percent agreement, PPV, or Sensitivity: $\geq 70\%$											
•Below threshold for relevant statistic type	0.25										
•Missing or different type of analysis provided	0.00										

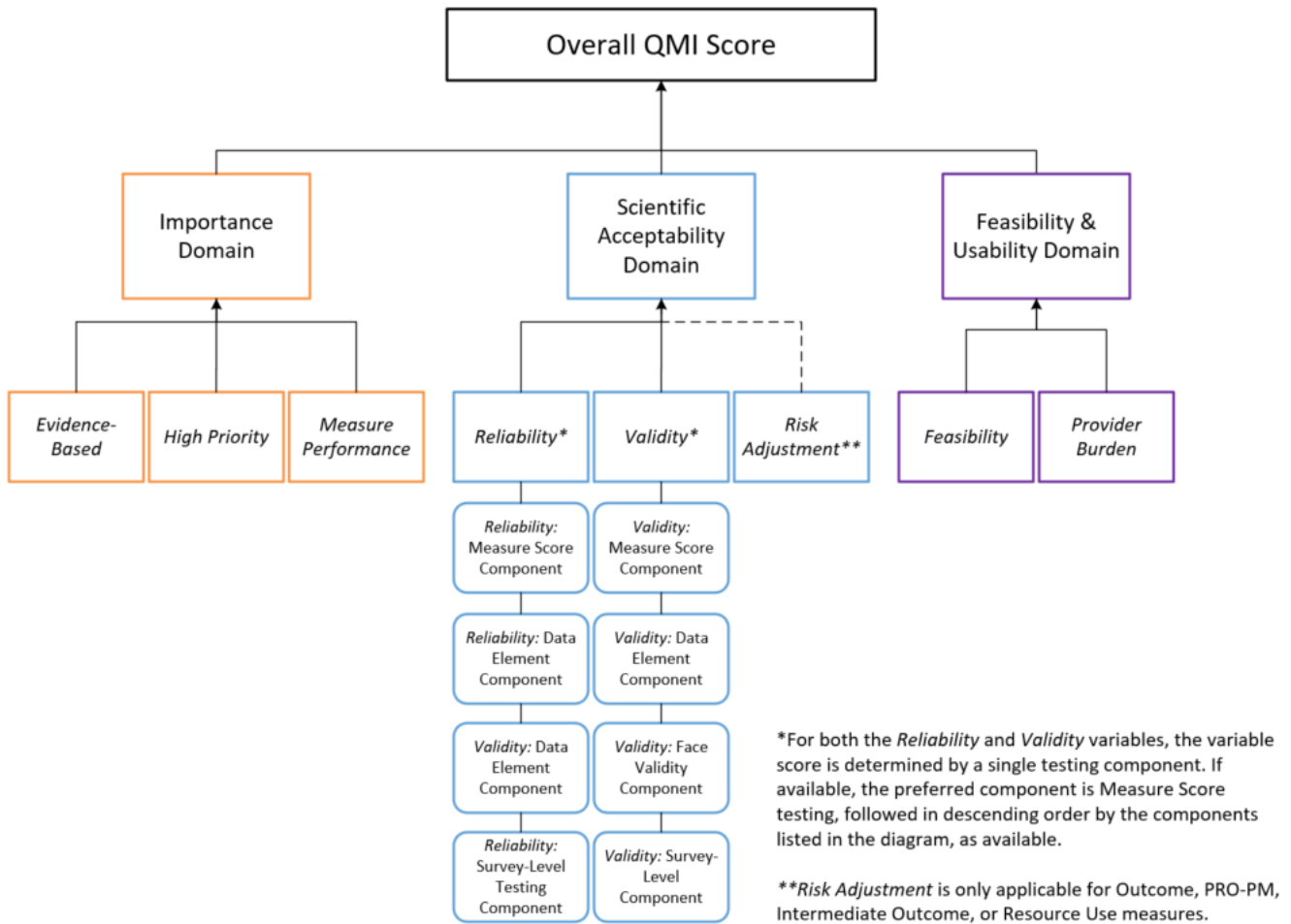
Variable Name	Scoring Approach Considerations and Limitations								
<p>Validity:</p> <p>Face Validity Component</p>	<p>Variable Scoring Approach: If neither empiric measure score validity testing nor data element validity are provided, the Scientific Acceptability domain score for <i>Validity</i> is based on the face validity assessment. Given that this is not an empirical test of measure score validity, the face validity scores are capped at 0.75 (acceptable), which is aligned with NQF’s approach of capping the validity score at moderate if this is the only information provided for a new measure.</p> <table border="1" data-bbox="381 604 1317 743"> <tr> <td>•Not applicable</td> <td>N/A</td> </tr> <tr> <td>•> 50% agreement that measure differentiates good from poor quality</td> <td>0.75</td> </tr> <tr> <td>• ≤50% agreement that measure differentiates good from poor quality</td> <td>0.25</td> </tr> <tr> <td>•Missing or vote on a different question</td> <td>0.00</td> </tr> </table> <p>Limitations: The QMI assumes the developer conducted the face validity assessment appropriately (e.g., transparent process, adequate number of experts, question framed appropriately, etc.) and only evaluates whether a majority of experts agreed that the measure can distinguish good from poor quality care.</p>	•Not applicable	N/A	•> 50% agreement that measure differentiates good from poor quality	0.75	• ≤50% agreement that measure differentiates good from poor quality	0.25	•Missing or vote on a different question	0.00
•Not applicable	N/A								
•> 50% agreement that measure differentiates good from poor quality	0.75								
• ≤50% agreement that measure differentiates good from poor quality	0.25								
•Missing or vote on a different question	0.00								
<p>Validity:</p> <p>Survey-Level Testing Component</p>	<p>Variable Scoring Approach: The Scientific Acceptability domain score for <i>Validity</i> is based on the survey-level testing component only if a measure reported survey-level testing results and did not provide measure score validity, data element validity or face validity testing results of the performance measure itself. The maximum survey-level testing score is capped at 0.75 (acceptable) to give lower priority to measures that only conducted survey-level testing and did not provide information on performance measure-level testing.</p> <table border="1" data-bbox="381 1136 1287 1274"> <tr> <td>•Not applicable</td> <td>N/A</td> </tr> <tr> <td>•Psychometric testing on survey or PROM conducted</td> <td>0.75</td> </tr> <tr> <td>•Not applicable</td> <td>N/A</td> </tr> <tr> <td>•Psychometric testing on survey or PROM not conducted</td> <td>0.00</td> </tr> </table> <p>Limitations: Current measure information documentation provided by NQF and CMS include survey-level testing as options in the validity field rather than in discrete fields. The QMI gives some credit when this information is provided in lieu of measure-level testing because developers may not have been aware that they should also provide performance measure testing results. The QMI also assumes the developer provided the survey-level testing that was appropriate, and the testing was applicable to the survey as specified in the performance measure. If more than one survey or PROM is included in the measure specifications, the QMI does not confirm that all surveys or PROMs have undergone psychometric testing. Therefore, the score for this variable should be interpreted as an initial indicator that does not negate the need for secondary review of a measure’s survey-level testing.</p>	•Not applicable	N/A	•Psychometric testing on survey or PROM conducted	0.75	•Not applicable	N/A	•Psychometric testing on survey or PROM not conducted	0.00
•Not applicable	N/A								
•Psychometric testing on survey or PROM conducted	0.75								
•Not applicable	N/A								
•Psychometric testing on survey or PROM not conducted	0.00								
<p>Risk Adjustment</p>	<p>Variable Scoring Approach: Measures should assess the need to risk adjust and/or risk stratify if they list the measure type as either outcome, PRO-PM, intermediate outcome, or resource use. This approach is aligned with the measure types that NQF identifies as having to demonstrate the need for and method of accounting for differences in the measured population. Other measure types are not scored on this variable and their Scientific Acceptability domain score is only based on the <i>Reliability</i> and <i>Validity</i> variables.</p>								

Variable Name	Scoring Approach Considerations and Limitations										
	<table border="1" data-bbox="381 388 1299 546"> <tr> <td>•Risk adjusted</td> <td>1.00</td> </tr> <tr> <td>•Not risk adjusted with rationale</td> <td></td> </tr> <tr> <td>•Not applicable</td> <td>N/A</td> </tr> <tr> <td>•Not applicable</td> <td>N/A</td> </tr> <tr> <td>•No risk adjustment or rationale provided</td> <td>0.00</td> </tr> </table> <p>Limitations: This variable only signifies whether the developer indicated the measure was risk-adjusted or risk-stratified and if not, whether there was a rationale provided. The QMI does not evaluate whether the rationale provided for choosing not to employ a risk adjustment strategy is acceptable (e.g., using exclusions in lieu of risk adjustment), nor whether the risk model is appropriately built and calibrated. Therefore, the score for this variable should be interpreted as an initial indicator that does not replace secondary review of a measure's strategy for risk adjustment.</p>	•Risk adjusted	1.00	•Not risk adjusted with rationale		•Not applicable	N/A	•Not applicable	N/A	•No risk adjustment or rationale provided	0.00
•Risk adjusted	1.00										
•Not risk adjusted with rationale											
•Not applicable	N/A										
•Not applicable	N/A										
•No risk adjustment or rationale provided	0.00										
Feasibility & Usability Domain											
Feasibility	<p>Variable Scoring Approach: This variable assesses the extent to which the data elements in the performance measure are accurate and consistently available for quality measurement.</p> <table border="1" data-bbox="381 955 1299 1092"> <tr> <td>•All data elements in electronically defined fields</td> <td>1.00</td> </tr> <tr> <td>•Some data elements in electronically defined fields</td> <td>0.75</td> </tr> <tr> <td>•No data elements in electronically defined fields</td> <td>0.25</td> </tr> <tr> <td>•Unable to determine format of data elements</td> <td>0.00</td> </tr> </table> <p>Limitations: The primary data source for this field are the NQF Submission Materials. Other measure information sources lack explicit information to allow for expert judgment to be reliably made. In future pre-rulemaking cycles, however, data for this field will be obtained directly from measure developers thus reducing missing data for this variable.</p>	•All data elements in electronically defined fields	1.00	•Some data elements in electronically defined fields	0.75	•No data elements in electronically defined fields	0.25	•Unable to determine format of data elements	0.00		
•All data elements in electronically defined fields	1.00										
•Some data elements in electronically defined fields	0.75										
•No data elements in electronically defined fields	0.25										
•Unable to determine format of data elements	0.00										
Provider Burden	<p>Variable Scoring Approach: This variable focuses on the measure calculation burden for the provider.</p> <table border="1" data-bbox="381 1344 1282 1575"> <tr> <td>•Claims •eCQM</td> <td>1.00</td> </tr> <tr> <td>•Registry •Claims + Quality data codes</td> <td>0.75</td> </tr> <tr> <td>•Other electronic abstraction method (e.g., CART)</td> <td></td> </tr> <tr> <td>•Manual abstraction</td> <td>0.25</td> </tr> <tr> <td>•Unable to determine calculation method</td> <td>0.00</td> </tr> </table> <p>Limitations: The forms used to score these measures for the QMI did not include specific questions on provider burden or calculation method. Therefore, this variable is based on a combination of information on submission method, data source, and intended implementation of the measure. Additionally, some measures are calculated using a hybrid approach so the QMI sought to give credit for the least burdensome calculation method wherever that was clearly noted in the forms. However, due to the heterogeneity in developer responses to the questions used to score this variable, some expert judgment was required.</p>	•Claims •eCQM	1.00	•Registry •Claims + Quality data codes	0.75	•Other electronic abstraction method (e.g., CART)		•Manual abstraction	0.25	•Unable to determine calculation method	0.00
•Claims •eCQM	1.00										
•Registry •Claims + Quality data codes	0.75										
•Other electronic abstraction method (e.g., CART)											
•Manual abstraction	0.25										
•Unable to determine calculation method	0.00										

Domain-Level and Overall Scoring




The scores of the variables within each domain are averaged and then multiplied by 100 to obtain the percentage for the domain score. Variables within each domain contribute equally to the individual overall domain score. The domain scores range from 0 to 100 with 0 representing the lowest possible domain score and 100 representing the highest possible domain score. Figure 1 provides more information on how each variable contributes to the domain level scores.

Figure 1. Domain Scoring Approach



The scores for the three domains are then averaged to obtain the overall QMI score for each measure. The overall scores range from 0 to 100, with 0 representing the lowest possible measure score and 100 representing the highest possible measure score. Figure 2 provides additional information on the overall scoring approach for the QMI.

Figure 2. QMI Scoring Summary

Domain Scores		QMI Score
 <p>Importance Domain Three variables scored: <i>Evidence, High Priority, Measure Performance</i></p>	<p>(Average variable scores x 100) Scores Range from 0-100</p>	<p>Average the Domain Scores x 100</p> <p>Scores range from 0-100</p>
 <p>Scientific Acceptability Domain Two to three variables scored depending on measure type: <i>Reliability, Validity, Risk Adjustment</i></p>	<p>(Average variable scores x 100) Scores Range from 0-100</p>	
 <p>Feasibility & Usability Domain Two variables scored: <i>Feasibility, Provider Burden</i></p>	<p>(Average variable scores x 100) Scores Range from 0-100</p>	

A factor to note is the number of variables is unequal among the three domains and across measure types. The Feasibility and Usability domain contains two variables, and the other two domains contain two or three variables depending on measure type. As a result, the variables have an unequal contribution to the overall QMI scores even though the domains have equal contribution.

Interpretation

A key purpose of the QMI is to highlight possible tradeoffs or risks within a measure to inform CMS decision making about measure implementation. Overall QMI scores are therefore best interpreted alongside variable-level scores.

In addition, measures that share similar classification characteristics are best compared to one another through stratification. For example, outcome measures do not need the same level of evidence as process measures; however, outcome measures must account for confounding factors through risk adjustment whereas process measures typically do not.

QMI Tool and Abstraction Guide

The QMI tool, developed in Microsoft Excel[®],ⁱⁱⁱ was utilized to capture measure-related data for assessment. Measure reviewers abstract measure data in the QMI tool to produce one record for each measure. These data were used to score the QMI variables and calculate a preliminary QMI score for each measure.

The tool opens a form with multiple tabs that groups the data elements and variables by domain. To standardize data entry, data elements were structured with drop-down menus and checkboxes, as allowable. Conditional logic was embedded into the form to reduce burden on measure reviewers. The tool’s design allowed measure reviewers to stop and save their progress and continue at any time.

ⁱⁱⁱ Microsoft 365 MSO (16.0.13127.21062) 64-bit

The team developed a QMI abstraction guide to provide instructions for navigating the QMI tool and completing tool fields. The abstraction guide was a resource for measure reviewers to identify appropriate measure information sources, determine specific locations for variable-specific data within an information source, and code responses appropriately. The fields in the tool and abstraction guide instructions can be incorporated into future pre-rulemaking cycles or Blueprint templates to clarify for developers which information is most pertinent and to standardize the format of information for CMS.

Limitations

Limitations of the QMI include the following:

- The QMI requires standardized measure information to calculate QMI variable scores. Variables have been defined to the extent feasible using current industry standards and acceptable thresholds. However, there may be instances where testing methodologies and thresholds outside of current standards need to be considered. In these instances, expert review would be required to determine the appropriateness of including these alternatives in the QMI.
- The field of measure development is quickly evolving. Although the QMI aims to align with current standards (e.g., CMS Blueprint v17.0² and the 2021 NQF measure evaluation criteria³), the integration of these standards into measure development practice can take time. To accommodate this, current variable scoring for measures is more lenient to allow broad application of the QMI to assess both existing and newly developed measures. It is expected that variable definitions and scoring will evolve over time to reflect evolving measure development standards.
- Values for certain QMI scoring variables are determined by a single best result provided by the measure developer, which may result in an upward bias of QMI scores. In these instances, supplemental qualitative expert review is required to confirm the appropriateness of the measure development and testing approach. For example, for the *Evidence-Based* variable, expert review is necessary to confirm evidence provided supports the measure as specified.
- In cases of complex testing results, more than one score is possible for a particular QMI variable. Currently, scores for these variables are determined by the best result available, which may result in an upward bias of QMI scores. Standardization of data requirements in the future will mitigate this issue by limiting information to key data elements required to assess the measure.
- Some QMI variables, such as *Risk Adjustment*, are limited in the scope of the assessment. For example, the QMI only identifies whether a measure is risk-adjusted or risk-stratified and if not, whether rationale is provided for not risk-adjusting or risk-stratifying. However, the QMI does not assess quality of the risk adjustment approach, such as assessing the discriminatory power or calibration of the model. It is anticipated that in the future, these variables may be further refined to provide more comprehensive information.

Next Steps

Next steps to enhance the QMI are to:

- Continue efforts to standardize the collection of data for measures submitted to CMS for consideration during pre-rulemaking and in other measure information sources. These efforts will reduce missing data and allow for more automated calculation of the QMI and potential future additions or enhancements to QMI variables.
- Continue to align the QMI with the CMS Blueprint and recommendations of the NQF Scientific Methods Panel (SMP) to ensure the QMI reflects current standards.^{2,3} For example, the assessment of *Reliability* in QMI is aligned with NQF SMP Draft Acceptable Reliability Thresholds (version 3.2) from the July 2021 SMP meeting.⁴ The team continues to follow the NQF SMP and will update the QMI where feasible.
- Explore more robust approaches to capturing the impact of a measure in areas such as health equity and estimated cost avoided. The team will examine revised data collection requirements to determine the feasibility of such variables.
- Continue to adapt the QMI for use in measures in various phases of development (e.g., conceptualization, specification), measure type (e.g., cost and resource use) and levels of analysis (e.g., health plan) to support broader use of the QMI.

Conclusion

Developed as a transparent and reliable scoring instrument based on standardized and objective definitions of quantifiable measure characteristics, the QMI can produce repeatable results yet is adaptable to evolving CMS priorities. The QMI benefits CMS and the measure development field by providing clear guidance for CMS expectations regarding quality measures developed for use in CMS programs and standardizes submitted measure information to ensure fair comparisons among measures. The QMI is broadly applicable across measures regardless of measure type, level of analysis, setting, and NQF endorsement status. Additionally, the QMI fulfills GAO recommendations by providing a systematic method to assess measures across reporting programs to ensure they align with CMS strategic objectives and meet industry standards. Finally, the QMI can be refined over time to reflect new CMS priorities and continued advancement in the field of measurement science.

References

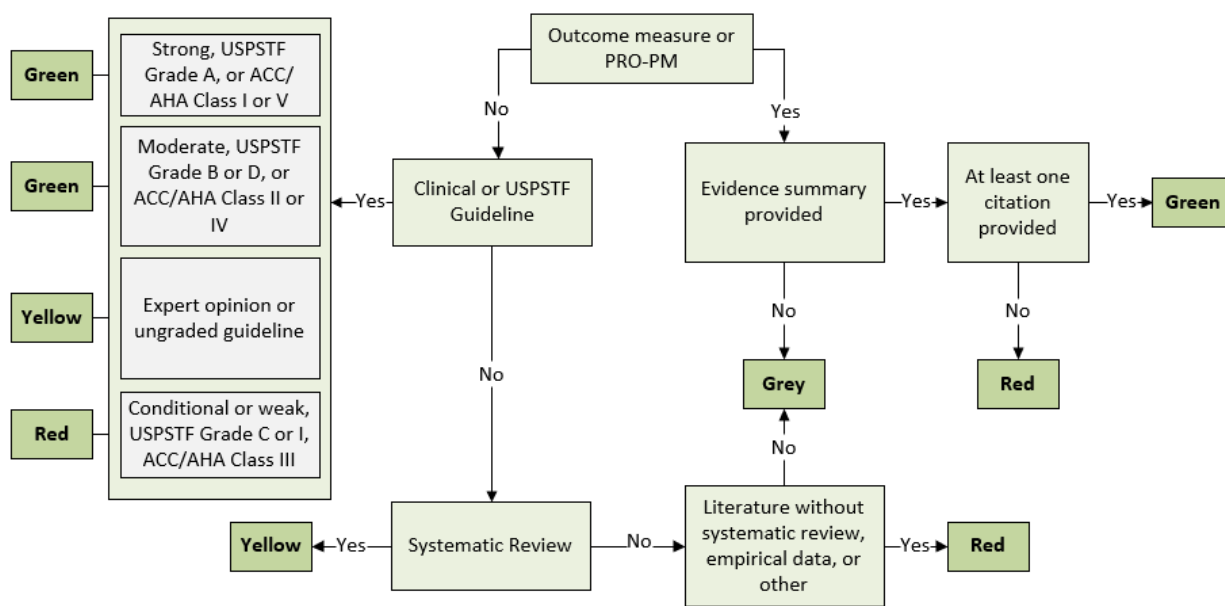
1. US Government Accountability Office. *Health Care Quality: CMS Could More Effectively Ensure Its Quality Measurement Activities Promote Its Objectives*. Washington, DC: US Government Accountability Office; 2019. <https://www.gao.gov/assets/710/701512.pdf>. Accessed January 10, 2022.
2. Centers for Medicare & Medicaid Services. *CMS Measures Management System Blueprint Version 17.0*. Baltimore, MD: US Department of Health and Human Services; 2021. Accessed January 20, 2022.
3. National Quality Forum. *Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement*. Washington, DC: National Quality Forum; 2021.
4. National Quality Forum. *Draft Acceptability Reliability Thresholds Version 3.2*. Washington, DC: National Quality Forum; 2021.

Appendix A – Variable Scoring Logic

Importance Domain

This section includes the scoring logic for the *Evidence Based*, *High Priority*, and *Measure Performance* variables.

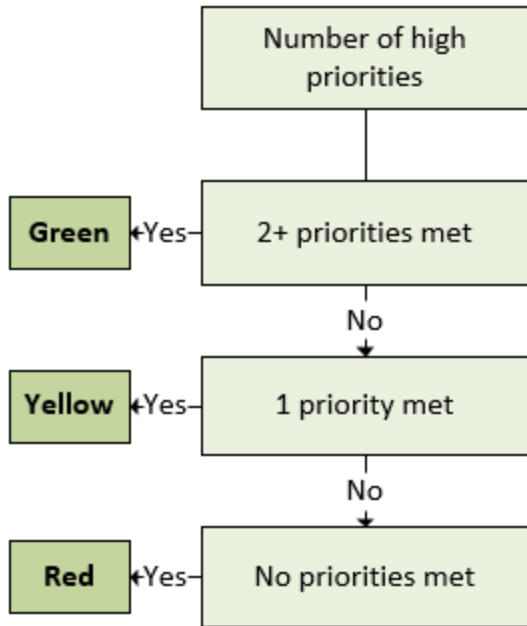
Evidence Based



Note: The highest rated piece of evidence provided by the developer is used for the QMI. If more than one guideline, choose the US-based guideline with the strongest recommendation.

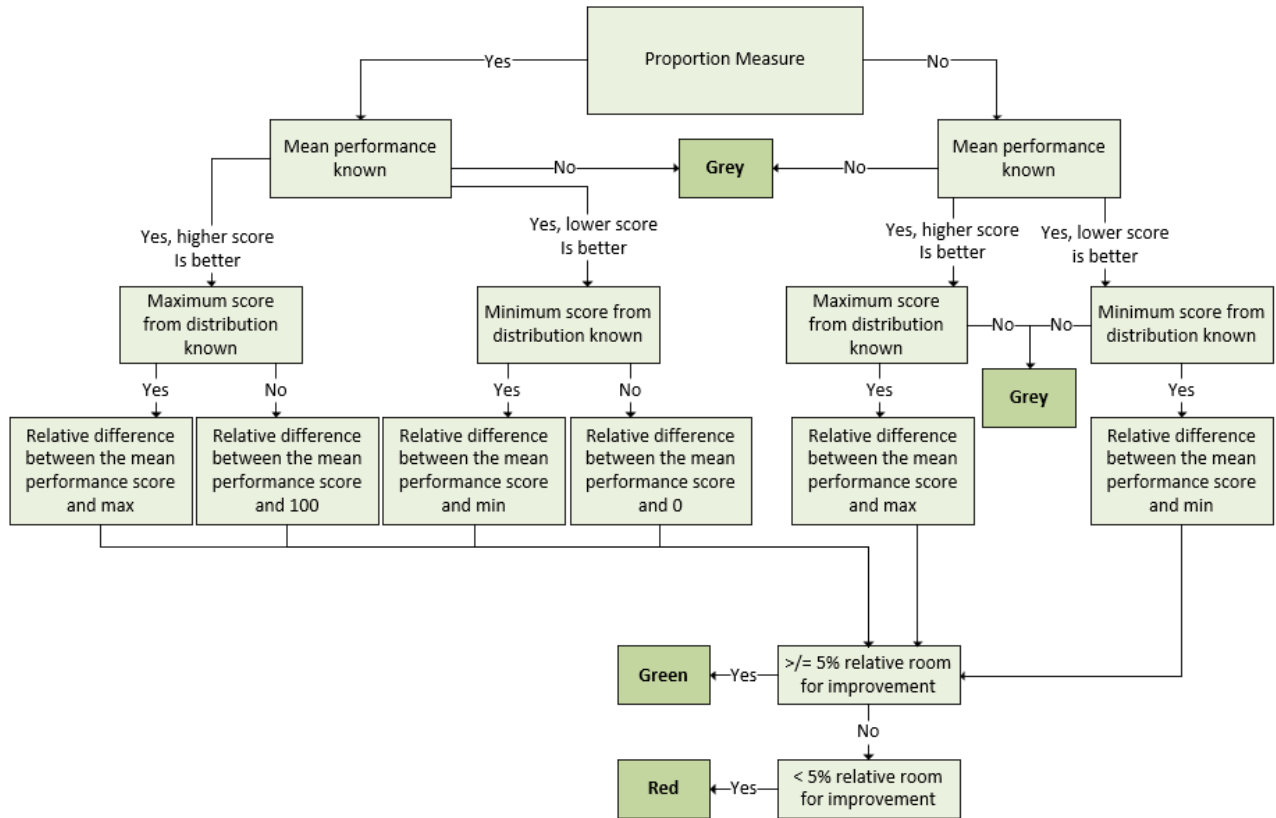
Note: If a measure is meant to disincentivize overuse or inappropriate use of a process, conditional or weak guidelines can be scored as green if they are aligned with the measure intent to reduce use of that service or process.

High Priority



Note: CMS priorities for this version of the QMI include PRO-PM, outcome, or intermediate outcome measures; digital measures defined as having at least one data source that is electronic; measures that selected "Equity" as the Meaningful Measures 1.0 area.

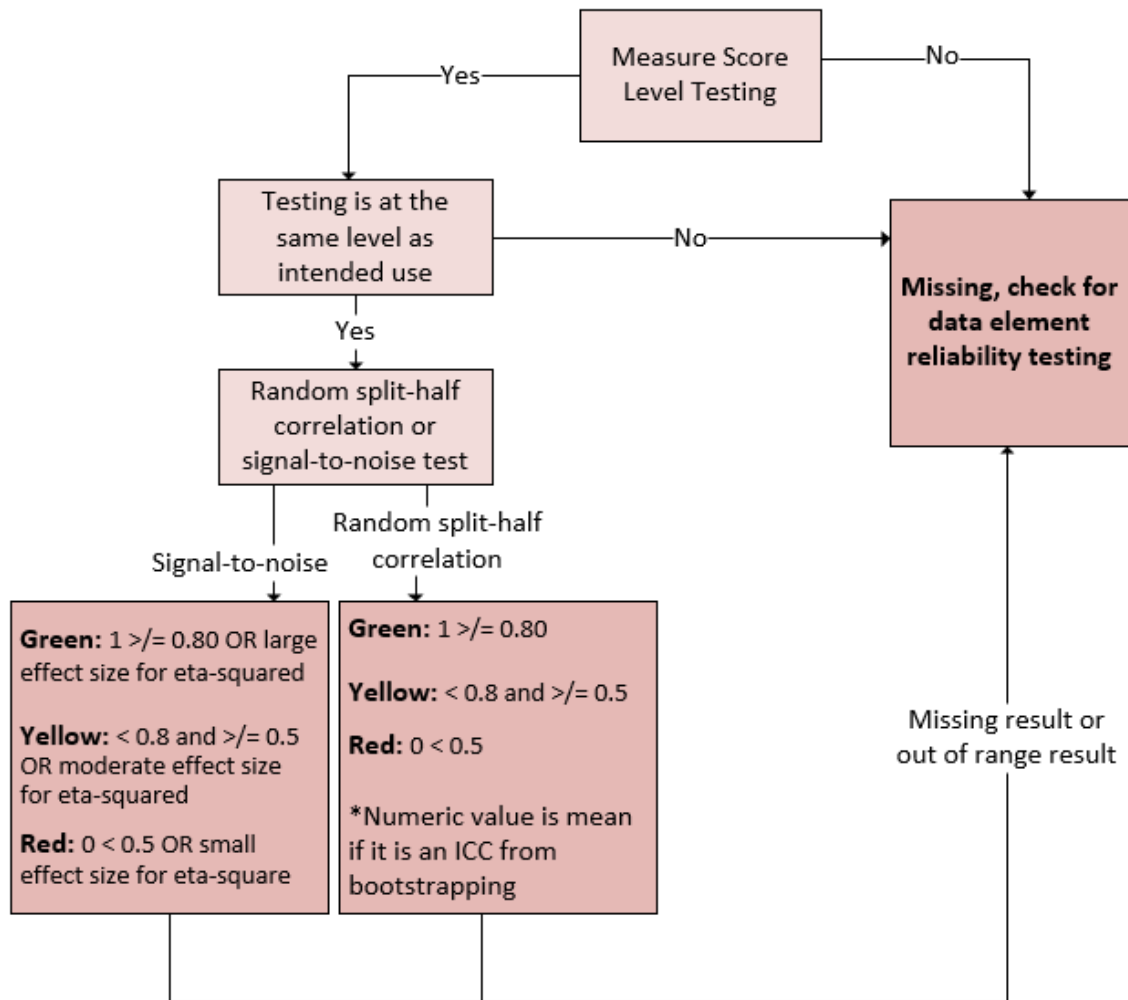
Measure Performance



Scientific Acceptability Domain

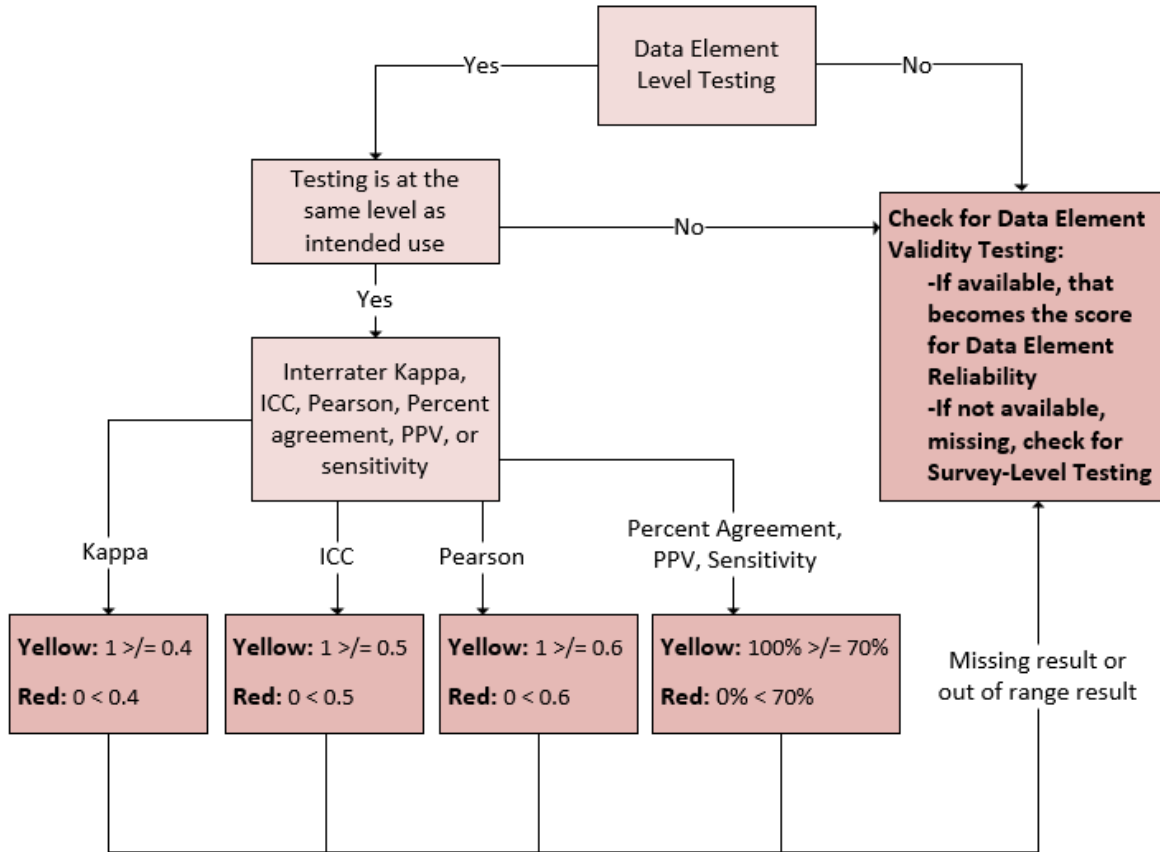
This section includes the scoring logic for the components of the *Reliability* variable: Measure Score Reliability, Data Element Reliability, Survey-Level Testing (*Reliability*); components of the *Validity* variable: Empiric Measure Score Validity, Data Element Validity, Face Validity, Survey-Level Testing (*Validity*); and *Risk Adjustment* variable.

Reliability – Measure Score Reliability



Note: When both results are provided, the result that produces the best Measure Score Reliability score is used to calculate the QMI score.

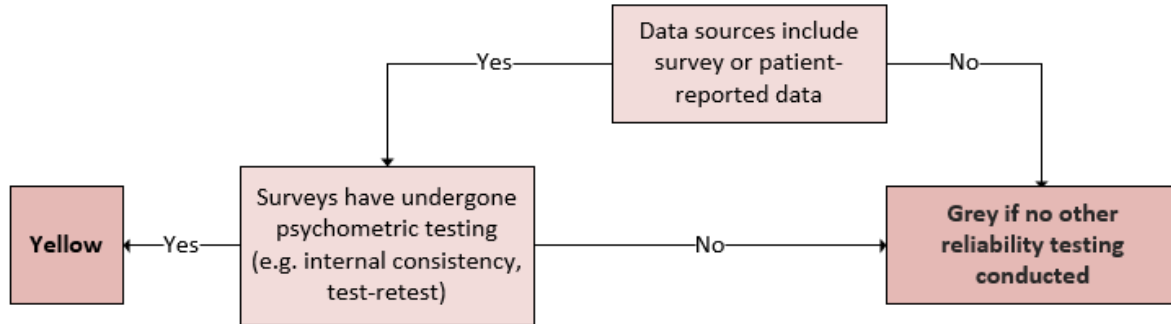
Reliability – Data Element Reliability



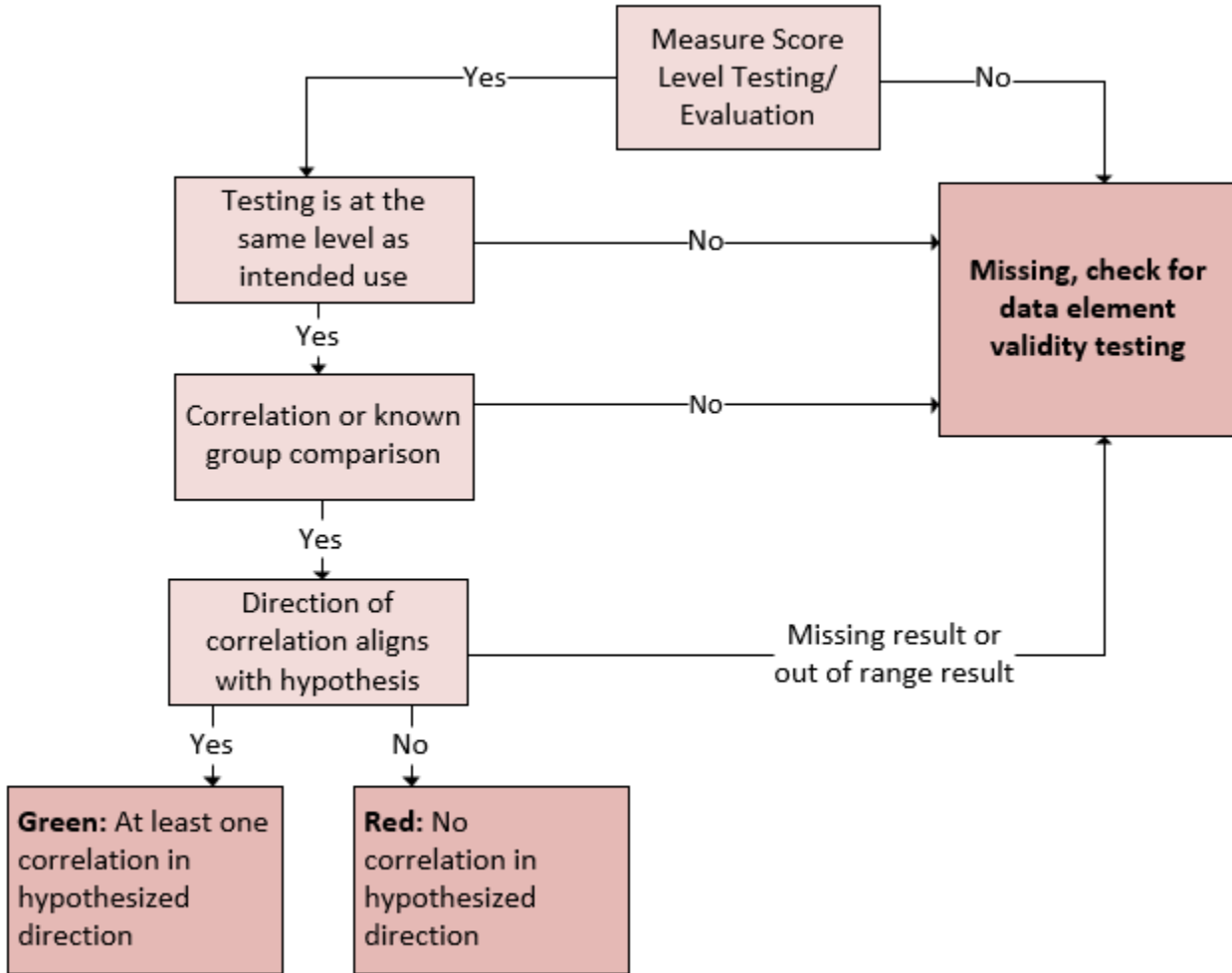
Note: When only summary level results are provided (e.g., denominator, numerator), the lowest result is used in the calculation of the QMI. If only one overall result is provided, that value is used.

Note: When more than one statistic is provided, the result that produces the best Data Element Reliability score is used in the calculation.

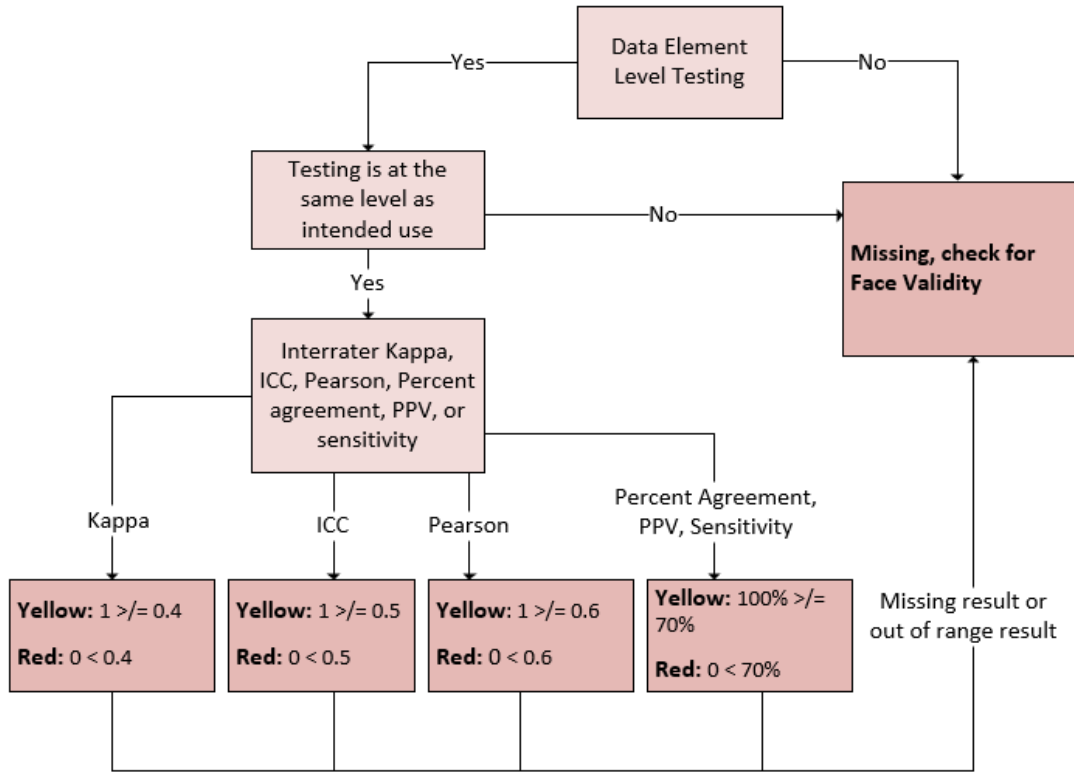
Reliability – Survey Level Testing



Validity – Empiric Measure Score Validity

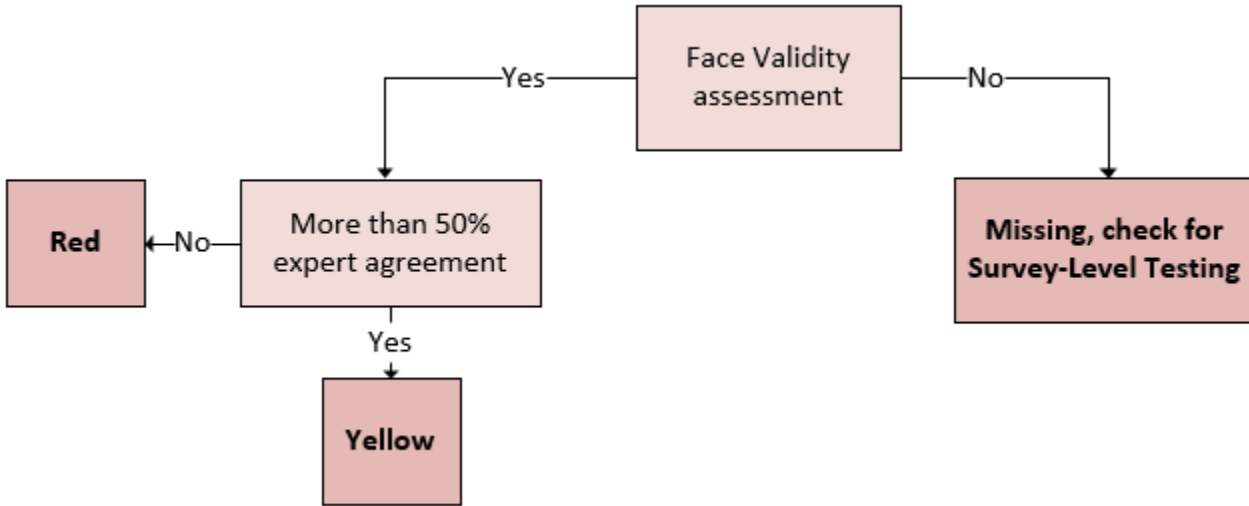


Validity – Data Element Validity

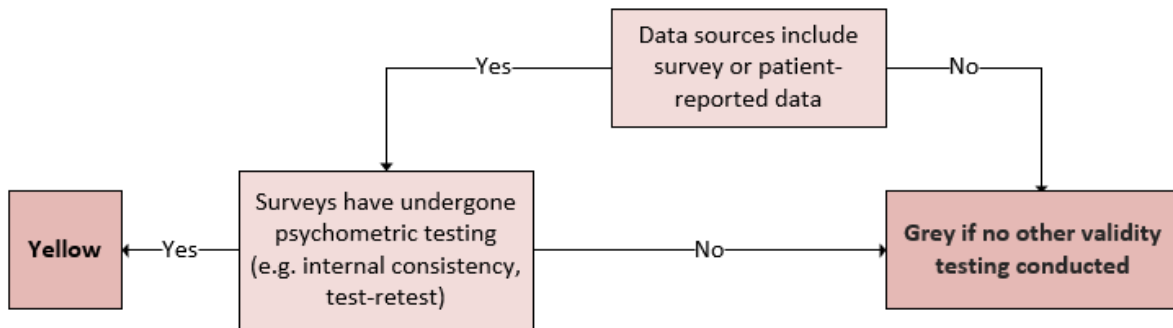


Note: The lowest data element value provided is used in the calculation of the QMI. When only summary level results are provided (e.g., denominator, numerator), the lowest is used to calculate the QMI. If only one overall result is provided, that value is used in the calculation.
 Note: When more than one statistic is provided, the result that produces the best Data Element Validity score is used.

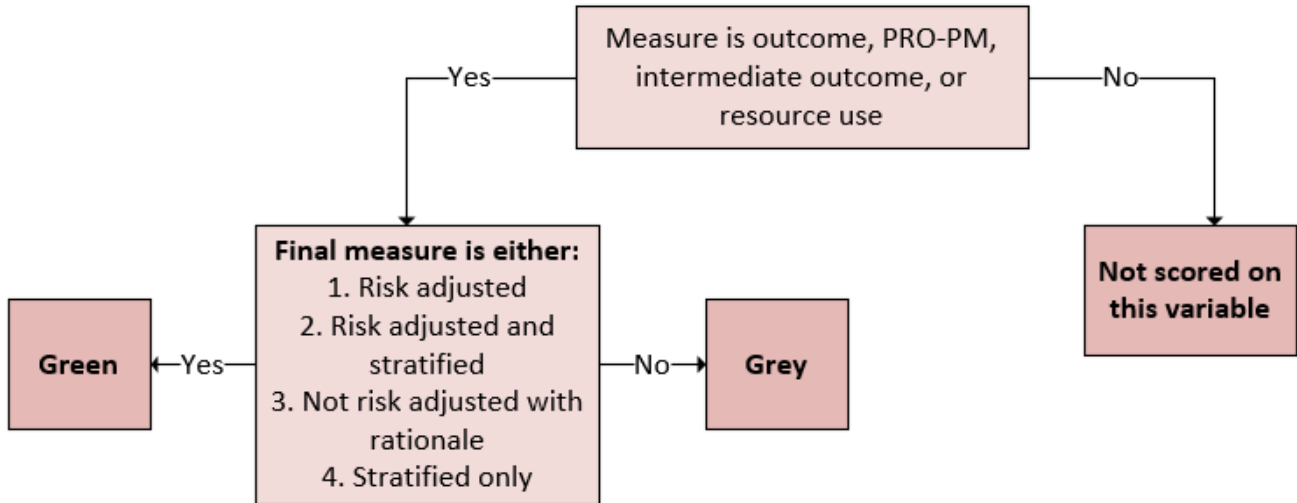
Validity – Face Validity



Validity – Survey-Level Testing



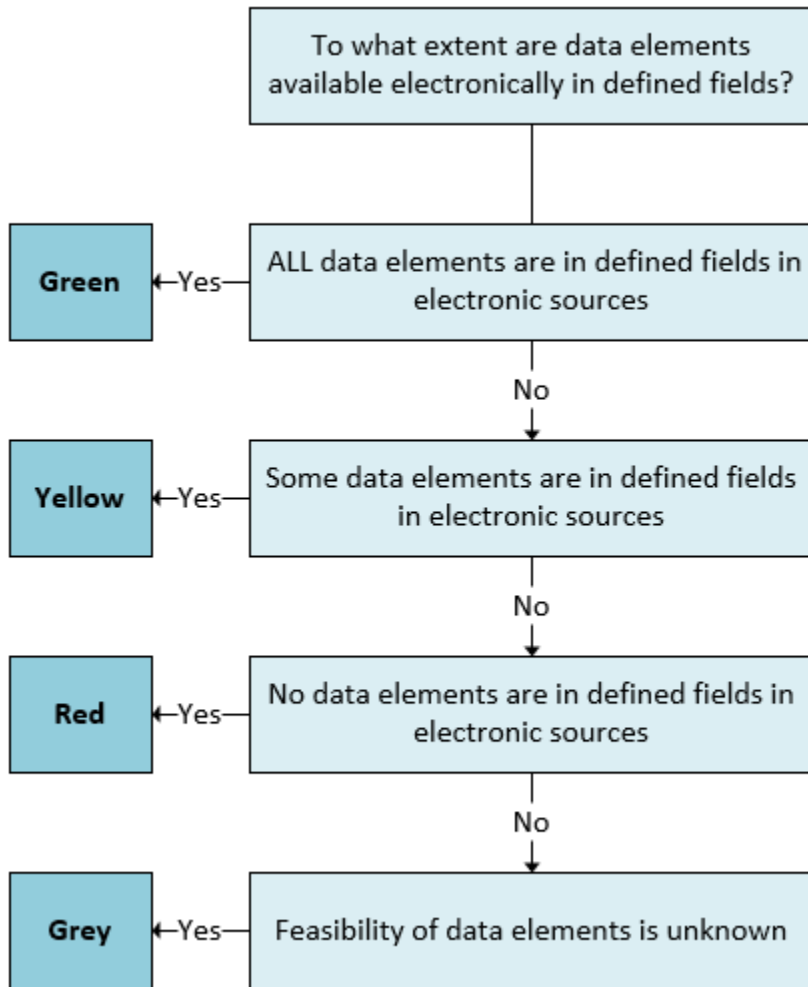
Risk Adjustment



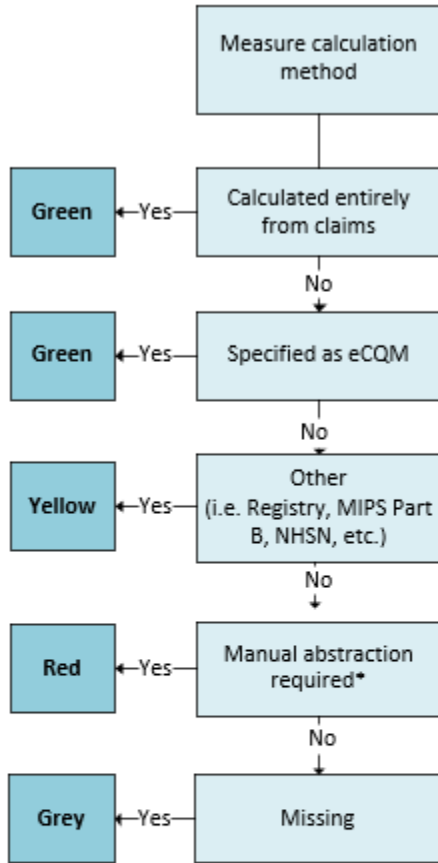
Feasibility and Usability Domain

This section includes the scoring logic for the *Feasibility* and *Provider Burden* variables.

Feasibility



Provider Burden



Note: If specified to allow for calculation using different methods, select the least burdensome method available to measured entities for measure calculation. If specified as a hybrid measure, select the most burdensome part of the hybrid specifications.

*In addition to requiring manual abstraction, data source is required to have 'Paper-record' and no 'Claims.'

Appendix B – Measure Scoring Example

Classification Variables

Variable	Measure Results
Meaningful Measure Healthcare Quality Priority	Promote Effective Communication & Coordination of Care
Meaningful Measure Quality Area	Admissions and Readmissions to Hospitals
Measure Type	Outcome
Composite Measure	No
Measure Submission Method	Claims
NQF Endorsement Status	Endorsed
Digital Measure	Yes
CQMC Core Measure Set	No

Importance Domain

Importance Domain Variables	Measure Response	Score
Evidence-Based: Outcome measures provided at least one citation for evidence and all other types of measures are supported by a clinical guideline, systematic review of the literature, and/or other evidence from the literature or empiric analyses	Outcome measure with at least one citation	1.00
High Priority: The measure's number of high priorities as defined by three CMS strategic measurement priorities: 1. Outcome, intermediate outcome, or PRO-PM 2. Digital 3. Equity	2 priorities (outcome, digital)	1.00
Measure Performance: The relative difference between the mean of measured entity performance scores and a performance benchmark	≥ 5% relative room for improvement	1.00
Total Domain Score (Average Variable Score x 100)		$(3.00 / 3) \times 100 = 100$

Scientific Acceptability Domain

Scientific Acceptability Domain Variables	Measure Response	Score
<p>Reliability: Testing component which provides the best possible score:</p> <ul style="list-style-type: none"> • Measure score reliability: Signal-to-noise or random split-half correlation results using measure scores for the same level of analysis as the intended measure use • Data element reliability: Interrater reliability results for each of the data elements required to score the measure • Survey-level testing reliability: Surveys and patient-reported outcome measures included in the performance measure have undergone psychometric testing 	Measure score reliability < 0.80 and ≥ 0.50	0.75
<p>Validity: Testing component which provides the best possible score:</p> <ul style="list-style-type: none"> • Empiric measure score validity: Empirical correlation using measure scores for the same types of measured entities as the intended measure use • Data element validity: Agreement with an authoritative source for each of the data elements required to score the measure • Face validity: Most experts consulted agree that the measure has face validity • Survey-level testing validity: Surveys and patient-reported outcome measures included in the performance measure have undergone psychometric testing 	Face validity had > 50% agreement that measure differentiates good from poor quality	0.75
<p>Risk Adjustment: Outcome, PRO-PM, intermediate outcome, and resource use measures are risk-adjusted</p>	Risk-adjusted	1.00
Total Domain Score (Average Variable Score x 100)		$(2.5 / 3) \times 100 = 83$

Feasibility and Usability Domain

Feasibility and Usability Domain Variables	Measure Response	Score
<p>Feasibility: To what extent are the specified data elements available electronically in defined fields (i.e., data elements needed to compute the performance measure score are in defined, computer-readable fields)</p>	All data elements in electronically defined fields	1.00
<p>Provider Burden: The least burdensome method available to providers to calculate measure scores each reporting period</p>	Claims	1.00
Total Domain Score (Average Variable Score x 100)		$(2.00 / 2) \times 100 = 100$

Overall QMI Score

