# **The Partnership for Quality Measurement (PQM) Measure Evaluation Rubric Worksheet**

**Purpose:** The purpose of this document is to provide guidance on the interpretation and application of the PQM Measure Evaluation Rubric (Rubric). The Rubric is used to evaluate clinical quality and cost/resource use measures that have been submitted to Battelle, a consensus-based entity (CBE), for PQM endorsement consideration.

**Intended Audience:** Reviewers of clinical quality and cost/resource performance measures for CBE-endorsement.

**Contents**

[The Partnership for Quality Measurement (PQM) Measure Evaluation Rubric Worksheet 1](#_Toc165890165)

[Overview 2](#_Toc165890166)

[Table 1a. Importance (pp. 1-3, PQM E&M Guidebook, 2023) 3](#_Toc165890167)

[Table 1b. Importance Rating 4](#_Toc165890168)

[Table 2a. Feasibility (p. 3, PQM E&M Guidebook, 2023) 5](#_Toc165890169)

[Table 2b. Feasibility Rating 6](#_Toc165890170)

[Table 3a.1. Scientific Acceptability (pp. 7-8, PQM E&M Guidebook, 2023) 7](#_Toc165890171)

[Table 3a.2. Scientific Acceptability (pp. 7-8, PQM E&M Guidebook, 2023) 8](#_Toc165890172)

[Table 3b. Scientific Acceptability Rating 9](#_Toc165890173)

[Table 4a. Equity (p. 8, PQM E&M Guidebook, 2023) 10](#_Toc165890174)

[Table 4b. Equity Rating 10](#_Toc165890175)

[Table 5a. Use and Usability (pp. 8-10, PQM E&M Guidebook, 2023) 11](#_Toc165890176)

[Table 5b. Use and Usability Rating 12](#_Toc165890177)

### **Overview**

The Rubric consists of five domains: **1.** [*Importance*](#_Table_1a._Importance), **2.** [*Feasibility*](#_Table_2a._Feasibility), **3.** [*Scientific Acceptability*](#_Table_3a.1._Scientific), **4.** [*Equity*](#_Table_4a._Equity), and **5.** [*Use and Usability*](#_Table_5a._Use).

The Rubric does not include must-pass criteria, nor algorithms for assigning a rating to each domain. Rather, the PQM Measure Evaluation Rubric guides reviewers to a rating of “Met”, “Not Met, but Addressable”, or “Not Met” based on the criteria listed for each domain ([PQM E&M Guidebook, 2023](https://p4qm.org/sites/default/files/2024-01/PQM-Measure-Evaluation-Rubric-v1.2.pdf)).

* **Met:** The measure information provided by the developer/steward meets the assertions/aspirations of the respective domain.
* **Not Met, but Addressable:** The measure information provided by the developer/steward does not meet the assertions/aspirations of the respective domain. However, can the measure developer/steward address any insufficiencies through reasonable changes[[1]](#footnote-2) to the measure (e.g., specifications, testing, evidence), which would improve its evaluation against the respective domain?
* **Not Met:** The measure information provided by the developer/steward does not meet the assertions/aspirations of the respective domain, and there are no reasonable changes to the measure (e.g., specifications, testing, evidence) that would allow the measure to meet the domain.

**Initial vs. Maintenance Endorsement:** In addition to new measures seeking initial (i.e., first) endorsement, previously endorsed measures undergo evaluation for maintenance of endorsement every five (5) years. The measure steward is responsible for making the necessary updates to the measure, informing E&M committees about any changes that are made to the measure, and providing the required measure information for the maintenance of endorsement evaluation. The Rubic and this Worksheet identify certain aspects of the measure submission that are required for ***[Initial endorsement]*** versus ***[Maintenance endorsement]***, in addition to what is required for both.

**Using this Worksheet:** This Worksheet identifies certain assertions and aspirations that measures should attempt to achieve for each of the 5 major domains of the Rubric. The credibility of the assertion is supported by evidence. Evidence in this context is any supporting information (e.g., testing results, research studies, technical expert panel recommendations) provided by the measure developer/steward that supports the assertions/aspirations. Measure reviewers (e.g., E&M committee members) express opinions and perspectives about the presented evidence and consider key questions to determine whether the measure assertions/aspirations made by the developer/steward are credible for meeting the respective domain.

### **Table 1a. Importance (**[**pp. 1-3**](https://p4qm.org/sites/default/files/2024-01/PQM-Measure-Evaluation-Rubric-v1.2.pdf)**, PQM E&M Guidebook, 2023)**

*Description: Extent to which the measure is important for making significant gains in health care quality or cost where there is variation in or overall, less-than-optimal performance. The measure focus is associated with a material outcome.*

| **Measure Assertion(s) / Aspiration(s)**  *There is an acceptable amount of certainty that:* | **Key Questions to Consider for Meeting the Domain** | **Potentially Relevant Submission Sections** | **Yes** | **No** | **Comments/Notes** |
| --- | --- | --- | --- | --- | --- |
| **1.** There is a clear business case for the measure ***[Initial endorsement]*** | * Do the anticipated impacts of the measure on outcomes justify the need for the measure and its use? In other words, will the existence and use of this measure lead to improved outcomes? | 1.1 Measure Rationale  2.1 Logic Model  2.3 Anticipated Impact |  |  |  |
| **2.** There is relationship between health care structures and/or processes and the desired outcome(s) | * Is the business case supported by credible evidence[[2]](#footnote-3) depicting a link between health care structures and/or processes to desired outcomes? For example, has the developer described health care services or procedures that can lead to improved outcomes?   + *[For structure, process, or intermediate outcome measures]* there is a demonstrated association between the measure focus and a material health outcome.   + *[For outcome and cost/resource use measures]* there is a demonstrated rationale for considering the measure focus, which is a material health outcome. This does not necessarily depend on a demonstration of or the degree of improvement. For example, the evidence may include a patient survey identifying care goals as important to decision making. * Does the evidence supporting the business case reflect the measure focus, the same population, and the same accountable entity? * If there are any limitations within the evidence, has this been made explicit? * Does the evidence include empirical studies of some kind? | 1.5 Measure Type  1.8 Level of Analysis  1.9 Care Setting  2.1 Logic Model  2.2 Evidence of Measure Importance |  |  |  |
| **3.** A gap in care exists, which this measure can identify and close | * ***[Initial endorsement]***If this is a new measure, does the developer indicate it is unique? If it is not unique, does the measure offer an advantage to existing measures and/or quality improvement programs? * ***[Maintenance endorsement]***If this measure is currently endorsed, does less than optimal performance and/or significant variation of the measure results still exist? | 2.4 Performance Gap  2.5 Health Care Quality Landscape |  |  |  |
| **4.** The patient population included in the measure find the measure focus meaningful. | * Is the focus of the measure valuable or meaningful to patients being captured in the measure? The developer may provide evidence from the literature or from patient focus groups/panels indicating patients being captured in the measure find the measure focus meaningful. | 2.6 Meaningfulness to Target Population |  |  |  |

### **Table 1b. Importance Rating**

*Review your assessment from Table 1a and assign your rating for this domain below. If you responded “Yes” to assertions 1-4 in Table 1a (or “Yes” to assertions 2-4 if a maintenance endorsement), your rating should be “Met”; if you responded “No” to at least one applicable assertion, your rating should be “Not Met, but Addressable” or “Not Met”.*

*Be sure to also include your rationale. If you assigned “Not Met, but Addressable,” what are the insufficiencies that can be reasonably be addressed to meet the requirements of this domain?*

* *Please note the following non-negotiables cannot be conditions: Lack of or unclear business case for the measure, lack of evidence supporting the business case, and (for maintenance measures) lack of sufficient evidence that a performance gap exists.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Met** | **Not Met, but Addressable** | **Not Met** | **Rationale** |
| Based on the evidence provided for the Importance domain, I assign the following rating and rationale: |  |  |  |  |

### **Table 2a. Feasibility (**[**p. 3**](https://p4qm.org/sites/default/files/2024-01/PQM-Measure-Evaluation-Rubric-v1.2.pdf)**, PQM E&M Guidebook, 2023)**

*Description: Extent to which the measure specifications (i.e., numerator, denominator, exclusions) require data that are readily available OR could be captured without undue burden AND can be implemented for performance measurement. There is an explicit articulation of the people, processes, and technology required for data collection and reporting.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Measure Assertion(s) / Aspiration(s)**  *There is an acceptable amount of certainty that:* | **Key Questions to Consider for Meeting the Domain** | **Potentially Relevant Submission Sections** | **Yes** | **No** | **Comments/Notes** |
| **1.** Required measure data are currently OR there is a near-term path for making measure data available from electronic sources (e.g., claims, electronic health records) ***[Initial endorsement]*** | * Did the developer conduct a feasibility assessment? If so, did the developer describe any data availability issues and how did these findings inform the measure specifications? * Are all required data elements routinely generated from digital or electronic sources (e.g., claims, electronic health records) and can promote interoperability? If not, has the developer described a near-term path to support routine data generation from electronic data sources? * ***[eCQMs only]*** If the measure is an electronic clinical quality measure (eCQM), did the developer complete the eCQM Feasibility Scorecard? If data availability issues were identified, did the developer describe a reasonable data collection strategy that can be implemented? | 1.20 Testing Data Sources  1.25 Data Sources  3.1 Feasibility Assessment  3.2 Feasibility Scorecard  3.3 Feasibility Informed Final Measure |  |  |  |
| **2.** Required measure data are routinely generated | * Are all required data elements routinely generated during normal care processes? | 3.1 Feasibility Assessment |  |  |  |
| **3.** Any proprietary components of the measure (e.g., codes, risk models, algorithms) do not cause substantial burden for the measure’s use. | * If the measure includes any proprietary information and/or has licensing fees, can the measure still be used without substantial burden? | 3.4 Proprietary Information |  |  |  |

### **Table 2b. Feasibility Rating**

*Review your assessment from Table 2a and assign your rating for this domain below. If you responded “Yes” to assertions 1 OR 2, AND “Yes” to assertion 3 in Table 2a, your rating should be “Met”; otherwise, your rating should be “Not Met, but Addressable” or “Not Met”.*

*Be sure to also include your rationale. If you assigned “Not Met, but Addressable,” what are the insufficiencies that can be reasonably be addressed to meet the requirements of this domain?*

* *Please note the following non-negotiables cannot be conditions: Significantly poor feasibility of the measure due to challenges with data availability or missingness and/or due to substantial proprietary mechanisms prohibiting a measure’s potential use.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Met** | **Not Met, but Addressable** | **Not Met** | **Rationale** |
| Based on the evidence provided for the Feasibility domain, I assign the following rating and rationale: |  |  |  |  |

### **Table 3a.1. Scientific Acceptability (**[**pp. 7-8**](https://p4qm.org/sites/default/files/2024-01/PQM-Measure-Evaluation-Rubric-v1.2_0.pdf)**, PQM E&M Guidebook, 2023)**

*Description: Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Measure Assertion(s) / Aspiration(s)** | **Key Questions to Consider for Meeting the Domain** | **Potentially Relevant Submission Sections** | **Yes** | **No** | **Comments/Notes** |
| ***Reliability***  *There is an acceptable amount of certainty that:* |  |  |  |  |  |
| **1.** The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability of care performance | * Can the measure be consistently implemented (i.e., are the measure specifications adequate)? If not, this may be due to unclear and/or highly complex measure specifications (including definitions, codes, data collection, and scoring) or very small sample sizes. | 1. Measure Specifications  4.1 Data and Samples |  |  |  |
| **2.** The measure data elements are repeatable, producing the same results when assessed in the same population under similar conditions in the same time period  AND/OR  The measure score is precise | * Has reliability testing been conducted for the measure, as specified? * If data element reliability testing was conducted for each level of accountability (e.g., clinician, hospital, health plan) for which the measure is specified, do the reliability testing results show all critical data elements (e.g., numerator, denominator, exclusions) are reproducible and meet accepted reliability thresholds (pp. [41-42](https://p4qm.org/sites/default/files/2023-10/Del-3-6-Endorsement-and-Maintenance-Guidebook-Final_0.pdf#page=44))? * If measure score reliability testing was conducted, for each level of accountability (e.g., clinician, hospital, health plan) for which the measure is specified, do the reliability testing results show the measure score is precise and meets accepted reliability thresholds (pp. [41-42](https://p4qm.org/sites/default/files/2023-10/Del-3-6-Endorsement-and-Maintenance-Guidebook-Final_0.pdf#page=44))? If not, does the developer propose a mitigation strategy for entities with reliability less than the threshold? * ***[Maintenance endorsement]***If no reliability testing was conducted, did the developer provide a reasonable rationale for why testing was not conducted or updated? | 1.8 Level of Analysis  4.2 Reliability |  |  |  |

### **Table 3a.2. Scientific Acceptability (**[**pp. 7-8**](https://p4qm.org/sites/default/files/2024-01/PQM-Measure-Evaluation-Rubric-v1.2_0.pdf)**, PQM E&M Guidebook, 2023)**

| **Measure Assertion(s) / Aspiration(s)** | **Key Questions to Consider for Meeting the Domain** | **Potentially Relevant Submission Sections** | **Yes** | **No** | **Comments/Notes** |
| --- | --- | --- | --- | --- | --- |
| ***Validity***  *There is an acceptable amount of certainty that:* |  |  |  |  |  |
| **3.** The measure’s data elements are correct | * If the developer conducted data element testing, are the measure’s critical data elements (e.g., numerator, denominator, exclusions) accurate/valid? If there are concerns, this may be due to missing or incorrect data or if there are multiple data sources not producing comparable results. | 1.8 Level of Analysis  4.3 Validity |  |  |  |
| **4.** The measure score correctly (i.e., without substantial bias or threats to validity) reflects the quality of care provided, adequately identifying good vs. poor quality care | * If the developer conducted accountable entity-level testing, does the measure score correctly reflect quality of care and distinguish good vs. poor quality of care? If there are concerns, this may be due to inaccuracy of the data elements (see above), inappropriate exclusions, and/or lack of appropriate risk adjustment and/or stratification (for outcome and cost/resource use measures). * ***[Initial endorsement]*** If face validity testing of the measure score was conducted, was the process planned and transparent?[[3]](#footnote-4) * If measure score validity testing was conducted, do the results indicate the measure score is valid indicator of quality and agree with what the developer predicted? If not, did the developer provide a reasonable rationale? * ***[Maintenance endorsement]***If no validity testing was conducted, did the developer provide reasonable rationale for why testing was not conducted or updated? * If an outcome or cost/resource use measure, was risk adjustment and/or stratification conducted, based on a well-defined conceptual model, AND does the risk adjustment analysis show:   + Variation in the prevalence of risk factors across providers AND   + Selected risk factors contribute to variation in the measured outcome AND   + Impact on provider scores at high or low extremes of risk due to presence of risk factors OR there is acceptable model performance | 4.3 Validity  4.4 Risk Adjustment |  |  |  |

### **Table 3b. Scientific Acceptability Rating**

*Review your assessment from Table 3a and assign your rating for this domain below. If you responded “Yes” to assertions 1-4 in Table 3a, your rating should be “Met”; if you responded “No” to at least one assertion, your rating should be “Not Met, but Addressable” or “Not Met”.*

*Be sure to also include your rationale. If you assigned “Not Met, but Addressable,” what are the insufficiencies that can be reasonably be addressed to meet the requirements of this domain?*

* *Please note the following non-negotiables cannot be conditions: Specifications, testing approach, results, or data descriptions are insufficient for the committee to apply the PQM Measure Evaluation Rubric and inappropriate methodology, calculations, formulas, or testing approach was used to demonstrate reliability or validity (this includes not testing the measure as specified).*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Met** | **Not Met, but Addressable** | **Not Met** | **Rationale** |
| Based on the evidence provided for the Scientific Acceptability domain, I assign the following rating and rationale: |  |  |  |  |

### **Table 4a. Equity (**[**p. 8**](https://p4qm.org/sites/default/files/2024-01/PQM-Measure-Evaluation-Rubric-v1.2_0.pdf)**, PQM E&M Guidebook, 2023)**

*Description: Extent to which the measure can identify differences in care for certain patient populations, which can be used to advance health equity and reduce disparities in care.*

***Note:*** *The Equity domain is optional, as Battelle recognizes some measures are not designed to advance health equity.[[4]](#footnote-5) Battelle continues to explore this, but to align with national priorities, Battelle encourages developers and stewards to address this domain, if and when possible.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Measure Assertion(s) / Aspiration(s)** | **Key Questions to Consider for Meeting the Domain** | **Potentially Relevant Submission Sections** | **Yes** | **No** | **Comments/Notes** |
| There is an acceptable amount of certainty that:  The measure contributes to efforts to address inequities in health care | * Does the developer show or describe statistical differences in the measure rates across different patient groups by their social and/or economic status? In other words, are there meaningful differences in the measure rates for patients of different races, geographic locations, incomes, etc.? | 5.1 Contributions Towards Advancing Equity  Consider also:  2.4 Performance Gap  4.4 Risk Adjustment |  |  |  |

### **Table 4b. Equity Rating**

*Review your assessment from Table 4a and assign your rating for this domain below. If you responded “Yes” to the assertion in Table 4a, your rating should be “Met”; if you responded “No”, your rating should be “Not Met, but Addressable” or “Not Met”.*

*Be sure to also include your rationale. If you assigned “Not Met, but Addressable,” what are the insufficiencies that can be reasonably be addressed to meet the requirements of this domain?*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Met** | **Not Met, but Addressable** | **Not Met** | **Rationale** |
| Based on the evidence provided for the Equity domain, I assign the following rating and rationale: |  |  |  |  |

### **Table 5a. Use and Usability (**[**pp. 8-10**](https://p4qm.org/sites/default/files/2024-01/PQM-Measure-Evaluation-Rubric-v1.2_0.pdf)**, PQM E&M Guidebook, 2023)**

*Description: Extent to which potential audiences (e.g., consumers, purchasers, providers, and policymakers) are using or could use measure results for both accountability and performance improvement to achieve the goal of high quality, efficient health care for individuals or populations.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Measure Assertion(s) / Aspiration(s)** | **Key Questions to Consider for Meeting the Domain** | **Potentially Relevant Submission Sections** | **Yes** | **No** | **Comments/Notes** |
| **Use** |  |  |  |  |  |
| The developer/steward attests that:  **1.** The measure is currently in use or there is a plan for use | * Has the developer indicated the measure is currently in use, or there is a plan for its use within a health care quality improvement program? This program can be a state-wide, regional, or national program or an internal quality improvement program, specific to an organization. | 6.1 Use |  |  |  |
| **Usability**  *There is an acceptable amount of certainty that:* |  |  |  |  |  |
| **2.** There is an explicit articulation of the resources and context that might facilitate or be a barrier to the way an accountable entity may improve on the measure | * Can the measure results be used by health care providers being held accountable to the measure to improve care? For example, if the measure holds hospitals accountable, can those hospitals use the measure results to establish actions for improving those measure results? * Are there any barriers the accountable entity can mitigate or facilitators the accountable entity could use to improve the measure results? | 6.2.1 Actions of Measured Entities to Improve Performance |  |  |  |
| **3.** There is a process for gathering feedback from measure users that can be used to improve the measure ***[Maintenance endorsement]*** | * Is there a process for gathering feedback from health care providers and others AND has the developer described how that feedback was taken into consideration to revise or update the measure? | 6.2.2 Feedback on Measure Performance  6.2.3 Consideration of Measure Feedback |  |  |  |
| **4.** If currently used, there is progress on improvement over time ***[Maintenance endorsement]*** | * If the measure is in use, has there been any improvement in the measure results over time or across patient groups (e.g., patients with social determinants of health)? * If the measure is in use, and there isn’t any improvement in measure results over time, has the developer provided an acceptable explanation as to why? | 6.2.4 Progress on Improvement |  |  |  |
| **5.** Any unexpected findings (positive or negative) do not outweigh the benefits of the measure ***[Maintenance endorsement]*** | * Are there any unexpected findings (positive or negative) AND if these findings negatively impact patients, does the developer describe how the benefits of the measure still outweigh these findings? | 6.2.5 Unexpected Findings |  |  |  |

### **Table 5b. Use and Usability Rating**

*Review your assessment from Table 5a and assign your rating for this domain below. If you responded “Yes” to assertions 1-5 in Table 5a (or “Yes” to assertions 1-2 if an initial endorsement), your rating should be “Met”; if you responded “No” to at least one applicable assertion, your rating should be “Not Met, but Addressable” or “Not Met”.*

*Be sure to also include your rationale. If you assigned “Not Met, but Addressable,” what are the insufficiencies that can be reasonably be addressed to meet the requirements of this domain?*

* *Please note the following non-negotiables that cannot be conditions: No plan for use within an accountability application (accountability applications are uses of measure performance results about identifiable, accountable entities to make judgments and decisions because of performance. For example, confidential reporting, reward, recognition, punishment, payment, or selection [e.g., public reporting, accreditation, performance-based payment, network inclusion/exclusion]).*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Met** | **Not Met, but Addressable** | **Not Met** | **Rationale** |
| Based on the evidence provided for the Use and Usability domain, I assign the following rating and rationale: |  |  |  |  |

1. When considering a reasonable change, consider whether the change or actions needed 1) can be completed within 3–5-year time frame, 2) can sufficiently and credibly support the measure assertion, and 3) are within the control of the measure developer/steward (e.g., data availability, resources) [↑](#footnote-ref-2)
2. Examples of credible evidence include systematic reviews, clinical practice guidelines, observational studies, and case studies. Should also include quality of evidence. [↑](#footnote-ref-3)
3. Face validity is accomplished through a systematic and transparent process, in which developers/stewards disclose identified relevant experts (e.g., clinicians, accountable entity representatives, those [patient, caregivers] with lived experience) and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed. [↑](#footnote-ref-4)
4. Health equity is the principle underlying a commitment to reduce—and eliminate—disparities in health and in its determinants, including social determinants. Health equity strives to ensure everyone has Any barriers or facilitators to whether the person or entity could use those ways are known and addressed (means/usability)a fair and just opportunity to be as healthy as possible. This requires removing obstacles to health, such as poverty, discrimination, and their consequences, including powerlessness and lack of access to good jobs with fair pay, quality education and housing, safe environments, and health. *(Culyer A. Equity - some theory and its policy implications. J Med Ethics. 2001;27(4):275-283).* [↑](#footnote-ref-5)