Measure Information and Justification Form and Instructions

***INSTRUCTIONS:*** *This form is primarily for measure developers to use as a guide when submitting measures. Measure developers may use information from the Measure Information and Justification Form (MIJF) for other purposes. CMS may ask measure developers to complete the MIJF for measures not submitted to the CMS consensus-based entity (CBE). Non-CMS Contracted Measure Developers or non-measure developers who elect to use the form for another purpose may edit the Project Overview section to reflect not having a measure development contract.*

*Please note: All CMS measure contract deliverables must meet accessibility standards as mandated in Section 508 of the Rehabilitation Act of 1973. This template is 508 compliant. You may not change the template format or non-italicized text. Any change could negatively impact 508 compliance and result in delays in the CMS review process. For guidance about 508 compliance, CMS’s* [*Creating Accessible Products*](https://www.cms.gov/es/node/1549751) *may be a helpful resource.*

*With approval from the Contracting Officer’s Representative (COR), measure developers may submit the CMS CBE submission form in lieu of the MIJF.*

***PLEASE DELETE THIS INTRODUCTORY SECTION (TEXT ABOVE THE LINE) AND REPLACE THE FORM-SPECIFIC REFERENCES ON THE LAST PAGE OF THE FORM WITH YOUR OWN REFERENCES BEFORE SUBMISSION. CMS REQUIRES NO SPECIFIC FORMAT FOR REFERENCES BUT BE COMPLETE AND CONSISTENT.***

***CMS-CONTRACTED MEASURE DEVELOPERS MUST USE THE MOST CURRENT PUBLISHED VERSION OF ALL TEMPLATES AND SHOULD CHECK THE*** [***CMS MMS HUB***](https://mmshub.cms.gov/) ***FOR UPDATES BEFORE SUBMISSION.***

**Project Title: *List the project title as it should appear in the web posting.***

**Date:**

Information included is current on *date*.

**Project Overview:**

The Centers for Medicare & Medicaid Services (CMS) contracted with *measure developer name* to develop *measure (set) name or description*. The *contract* name is *insert name*. The *contract* number is *project number*.

 **Measure Name/Title**

*The name should be brief and include the measure focus and the target population.*

1. **Descriptive Information**

1.1 Measure Type

*Identify a measure type from the list. Patient-reported outcome-based performance measures (PRO-PMs) include health-related quality of life, functional status, symptom burden, and health-related behavior. For composite measures, please also identify the measure type of the components.*

☐process

☐outcome

☐PRO-PM

☐cost /resource use

☐efficiency

☐structure

☐intermediate outcome

☐population health

☐composite

☐process

☐outcome

☐other

☐other

1.2 Brief Description of Measure

*This description should be concise and include type of score, measure focus, target population, and time frame.*

1.3 If Paired or Grouped

*Provide the reason why you must report the measure with other measures to interpret results appropriately.*

1. [**Measure Specifications**](https://mmshub.cms.gov/measure-lifecycle/measure-specification/overview)****

*These items follow the CMS requirements for measure submission and provide information required for measure evaluation.*

2.1 Measure-Specific Webpage

*Provide a Uniform Resource Locator (URL) link, if available, to a webpage where you can obtain current, detailed specifications, including code lists, risk adjustment model details, and supplemental materials. Do not enter a URL linking to a home page or to general information. If no URL is available, indicate N/A.*

2.2 If this is an electronic clinical quality measure (eCQM)

*If not an eCQM, state N/A.*

*If an eCQM, attach the zipped output from the Measure Authoring Tool (MAT*) *and Bonnie testing results, when testing complete. Use the specification fields from the online form for the plain language description of the specifications.*

2.3 Data Dictionary, Code Table, or Value Sets*.*

*Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). The preferred file format is either .xls or .csv. If not used, contact CMS for further directions.*

2.4 For an instrument-based measure

*If not an instrument-based measure, indicate N/A.*

*Attach copy of the instrument, if available*

*Indicate the responder (i.e., patient, family or other caregiver, clinician).*

2.5 Updates since last submission

*If this is the first submission, state N/A.*

*Are there changes to the specifications since the last updates/submission? If yes, update the specifications in 2.1 and 2.6-2.24 and explain the reasons for the changes.*

*Briefly describe any changes to the measure specifications since the last endorsement date, if CBE-endorsed, and explain the reasons for the changes.*

2.6 Numerator Statement

*Briefly describe the measure focus or what the measure measures about the target population—cases from the target population with the target process, condition, or event based on the evidence.*

*For example:*

*Patients in the target population who received/had [measure focus] {during [time frame]} if different from the target population.*

*Do not include the rationale for the measure.*

*For outcome measures, state the measured outcome. Describe calculation of the risk-adjusted outcome later in the calculation algorithm.*

2.7 Numerator Details

*Include all information necessary to identify and calculate the cases from the target population with the target process, condition, event, or outcome. Provide definitions and specific data collection items and responses. For measures based on a coded data set, identify the code set, the specific codes, and the code descriptors**. If the list of codes and descriptors exceeds one page, provide the list in a Microsoft Excel or .csv file in the format listed in 2.3.*

*For outcome measures, describe how to identify and count the observed outcome. The calculation algorithm should also describe how to calculate the risk adjustment.*

*Provide the time period for measure data aggregation (e.g., 12 months, 3 years, another specified look-back period).*

2.8 Denominator Statement

*Provide a narrative description of the broadest population (based on the evidence) for which the target process, condition, event, or outcome is applicable. Include the time period for measure data aggregation, if different from the numerator.*

*Example*

*Patient [age] with [condition] in [setting] during [time frame]*

*For outcome measures, state the target population for the outcome. The calculation algorithm should also describe how to calculate the risk adjustment.*

2.9 Denominator Details

*Provide all definitions and instructions needed to identify and calculate the target population/ denominator (e.g., definitions, time period for data collection, specific data collection items/responses, codes/value sets). For measures based on a coded data set, identify the code set, the specific codes, descriptors, definitions, and specific data collection items as appropriate. (If the list of codes and descriptors exceeds one page, provide the list in an .xls or .csv file in the format listed in 2.3.)*

*For outcome measures, describe how to identify the target population. The calculation algorithm should also describe how to calculate the risk adjustment.*

2.10 Denominator Exclusions

*If no denominator exclusions, state N/A and skip to 2.12.*

*Identify patients in the target population who should not receive the process (i.e., medical treatment), or are not eligible for the outcome for some other reason, particularly if their inclusion may bias results. Exclusions should be evidence-based. If no denominator exclusions, indicate N/A and skip to 3.12.*

*Example*

*Patients in the [target population] who [have some additional characteristic, condition, procedure]*

2.11 Denominator Exclusion Details

*Provide all information needed to identify and calculate exclusions from the denominator (e.g., definitions and/or specific data collection items and responses). For measures based on a coded data set, identify the code set, specific codes, descriptors, definitions, and specific data collection items for the codes as appropriate. Provide lists of individual codes with descriptors exceeding one page in an .xls or .csv file in the required format listed in 2.3.*

2.12 Type of Score

☐count

☐rate/proportion

☐ratio

☐categorical (e.g., yes or no)

☐continuous variable (CV) (e.g., an average)

☐composite/scale

☐other (specify) Click or tap here to enter text.

2.13 Interpretation of Score

*Provide an interpretation classifying whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score.*

2.14 Risk Adjustment Type

*Select the risk adjustment type. Provide specifications for risk stratification in 2.15 and for the statistical model in 2.16-2.17.*

☐no risk adjustment or risk stratification

☐stratification by risk category/subgroup

☐statistical risk model

☐other

2.15 Stratification Details/Variables

*Provide instructions for calculating the measure by category (e.g., age), including the stratification variables, all codes, logic, definitions, specific data collection items/responses, and the risk model covariate and coefficients for the clinically adjusted version of the measure, when appropriate. Provide lists of individual codes with descriptors exceeding one page in an .xls or .csv file in the required format listed in 2.3.*

2.16 Calculation Algorithm/Measure Logic

*Describe the sequence of steps necessary to calculate the measure score, including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; time period of data; and any other calculations.*

*You may provide a diagram of the calculation algorithm/measure logic at a measure-specific webpage URL identified in 2.1 or in an appendix.*

2.17 Sampling

*If the measure is based on a sample or survey, provide instructions for obtaining the sample and conducting the survey, along with minimum response rate required. If the measure is an instrument-based quality measure (e.g., PRO-PM), identify whether (and how) to allow proxy responses.*

2.18 Survey/Patient-Reported Data

*If the measure is not based on a survey or instrument, state N/A.*

*If the measure is based on a survey or instrument, provide instructions for data collection and guidance on the minimum response rate. If the measure is a PRO-PM, specify how to report the calculation of response rates with quality measure results.*

*Specify how to handle missing data (e.g., imputation, delete case). This item is a requirement for composite measures and PRO-PMs.*

2.19 Data Source

*Indicate all sources for which the measure developer specified and tested. Provide testing for all sources of data specified and intended for measure implementation. If using different data sources for the numerator and denominator, indicate “numerator” or “denominator” with each source.*

☐administrative data

☐claims data

☐paper patient medical records

☐electronic patient medical records

☐electronic clinical data

☐registries

☐standardized patient assessments

☐patient-reported data and surveys

☐non-medical data

☐other—describe in 2.20

2.20 Data Source or Collection Instrument

*Identify the specific data source/data collection instrument (e.g., name of database, clinical registry, collection instrument). If the measure is instrument-based (e.g., PRO-PM), identify the specific tools/instruments used to collect the measure information and standard methods, modes, and languages of administration.*

2.21 Data Source or Collection Instrument (Reference)

*Provide the reference for the data source or collection instrument. Attach a copy or specify the URL.*

2.22 Level of Analysis

*Indicate only the levels for which the measure developer specified and tested.*

☐individual clinician

☐group/practice

☐hospital/facility/agency

☐health plan

☐accountable care organization

☐geographic population

☐other (specify) Click or tap here to enter text.

2.23 Care Setting

*Indicate only the settings for which the measure developer specified and tested.*

☐ambulatory surgery center

☐clinician office/clinic

☐outpatient rehabilitation

☐urgent care – ambulatory

☐behavioral health: inpatient

☐behavioral health: outpatient

☐dialysis facility

☐emergency medical services/ambulance

☐emergency department

☐home health

☐hospice

☐hospital

☐hospital: critical care

☐hospital: acute care facility

☐imaging facility

☐laboratory

☐pharmacy

☐nursing home/skilled nursing facility (SNF)

☐inpatient rehabilitation facility (IRF)

☐long-term acute care

☐birthing center

☐no applicable care setting

☐other (specify) Click or tap here to enter text.

2.24 Composite Measure

*This section is for additional specifications as needed. Use it for aggregation and weighting rules or calculation of individual quality measures.*

1. [**Importance**](https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/importance)****

3.1 Evidence to Support the Measure Focus (for reference only)

*The measure focus is evidence-based, demonstrated as*

* *a health outcome with a rationale supporting the relationship of the health outcome to processes or structures of care*
* *an intermediate outcome with a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence the measured intermediate outcome leads to a desired health outcome*
* *a patient-reported measure with evidence the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information, or the patient experience with care is correlated with desired outcomes*
* *efficiency measure with evidence for the quality component implied in experience with care; measures of efficiency combine the concepts of resource use and quality*

*Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events compared with zero are appropriate outcomes for public reporting and quality improvement.*

*The preferred systems for grading the evidence are the* [*United States Preventive Services Task Force (USPSTF)*](https://www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions) *grading definitions and methods, or* [*Grading of Recommendation, Assessment, Development, and Evaluation (GRADE)*](http://www.gradeworkinggroup.org/) *guidelines.*

3.1.1 Logic Model

*Briefly state or diagram the steps between the health care structures and processes (e.g., interventions, services) and the patient’s health outcome(s). General, non-technical audiences should easily understand the relationships in the diagram. Indicate the structure, process, or outcome for measurement.*

3.1.2 Value and Meaningfulness

*If this is a patient-reported measure, provide evidence the target population values the measured outcome, process, or structure and finds it meaningful. Describe how and from whom you obtained input.*

*\*\*RESPOND TO ONLY ONE OF THE NEXT THREE SECTIONS - EITHER 3.1.3, 3.1.4, or 3.1.5 \*\**

3.1.3 Empirical Data (for outcome measures) – as applicable

*Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one health care structure, process, intervention, or service.*

3.1.4 Systematic Review of the Evidence (for intermediate outcome, process, or structure quality measures, include those that are instrument-based) – as applicable

*What is the source of the systematic review of the body of evidence supporting the quality measure? A systematic review is a scientific investigation focusing on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar, but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (*[*Institute of Medicine, 2011*](https://www.nap.edu/catalog/13059/finding-what-works-in-health-care-standards-for-systematic-reviews)*)*

☐Clinical Practice Guideline recommendation (with evidence review)

☐USPSTF recommendation

☐other systematic review and grading of the body of evidence (e.g., Cochrane Collaboration, [AHRQ Evidence-based Practice Centers](https://effectivehealthcare.ahrq.gov/about/epc))

☐other

*For each systematic review, populate the table. Make as many copies of the table as needed to accommodate each systematic review.*

| **Source of Systematic Review (SR)*** **Title**
* **Author**
* **Date**
* **Citation, including page number**
* **Uniform Resource Locator (URL)**
 |  |
| --- | --- |
| Quote the guideline or recommendation verbatim about the process, structure, or intermediate outcome for measurement. If not a guideline, summarize the conclusions from the SR. |  |
| Grade assigned to the **evidence** associated with the recommendation with the definition of the grade. |  |
| Provide all other grades and definitions from the evidence grading system. |  |
| Grade assigned to the **recommendation** with definition of the grade. |  |
| Provide all other grades and definitions from the recommendation grading system. |  |
| Body of evidence* Quantity – how many studies?
* Quality – what types of studies?
 |  |
| Estimates of benefit and consistency across studies.  |  |
| What were the harms identified? |  |
| Identify any new studies conducted since the initial SR. Do the new studies change the conclusions from the initial SR?  |  |

3.1.5 Other Source of Evidence – as applicable

*If source of evidence is not from a clinical practice guideline, USPSTF, or SR, describe the evidence on which quality measure is based.*

3.1.5.1 Briefly Synthesize the Evidence

*A list of references without a summary is not acceptable.*

3.1.5.2 Process Used to Identify the Evidence

*Identify guideline recommendation number and/or page number and quote verbatim the specific guideline recommendation.*

3.1.5.3 Citation(s) for the Evidence

*Grade assigned to the quoted recommendation with definition of the grade.*

3.2 Performance Gap – Opportunity for Improvement

3.2.1 Rationale

*Briefly explain the rationale for this measure (i.e., benefits or improvements in quality envisioned by use of this measure).*

*If the measure is a composite, a combination of component measure scores, all-or-none, or any-or-none, describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually. Describe the area of quality measured, component measures, and the relationship of the component measure to the overall composite and to each other (whether reflective or formative model used to develop this measure, and whether components are correlated. Describe how the aggregation and weighting of the components measures are consistent with the stated quality construct and rationale.*

3.2.2 Performance Scores

*Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. Include the mean, standard deviation, minimum, maximum, interquartile range, and scores by decile. Describe the data source, including number of measured entities, number of patients, dates of data, and, if a sample, characteristics that the entities include. Also use this information to address the subcriterion on improvement in 6.1.1.*

3.2.3 Summary of Data Indicating Opportunity

*If there is no or limited performance data in 3.2.2 on the measure as specified, provide a summary of data from the literature indicating opportunity for improvement or overall, less-than-optimal performance on the specific focus of measurement. Include citations.*

3.2.4 Equity/Disparities

*Provide data on how the measure, as specified, addresses disparities—current and over time—by population group (e.g., race or ethnicity, gender, age, insurance status, socioeconomic factors, and disability). Describe the data source, including number of measured entities, number of patients, and dates of the data. If the data are from a sample, include characteristics of the entities. What is the impacted demographic and what component is the measure addressing and how? For measures showing high levels of performance (i.e., topped out), disparities data may demonstrate an opportunity for improvement/gap in care for certain subpopulations. Also use this information to address the subcriterion on improvement in 6.1.1.*

3.2.5 If no or limited equity/disparities data, provide summary of data.

*If there are no or limited data on disparities reported from the measure as specified in 3.2.2, provide a summary of data from the literature addressing disparities in care on the specific focus of measurement and include citations. The summary is not necessary if you provided performance data in 3.2.2.*

3.3 [Harmonization](https://mmshub.cms.gov/measure-lifecycle/measure-specification/harmonization)

*If this measure conceptually addresses either the same measure focus or the same target population as other measure(s), are the measure specifications harmonized to the extent possible?*

*If there is not complete harmonization of the measure specifications, identify the differences, rationale, and impact on interpretability and data collection burden.*

3.3.1 Related and Competing Measures

*If a measure meets other criteria and there are related measures (either the same measure focus or target population) or competing measures (both the same measure focus and same target population), you compared the measures to address harmonization and/or selection of the best measure.*

3.3.1.1 Relation to Other Measures

Are there related measures or competing measures?

☐yes

☐no

*If there are related measures (i.e., conceptually related by same measure focus or same target population) or competing measures (i.e., same measure focus and same target population), , list the title of all related and/or competing measures and list the CMS CBE number, if applicable.*

3.3.1.2 Competing Measures

*If this measure conceptually addresses the same measure focus and the same target population as other measure(s), describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality), or provide a rationale for the additive value of endorsing an additional measure. Provide analyses when possible.*

1. [**Scientific Acceptability**](https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/scientific-acceptability)****

4.1 What is the Source of Data Used for Testing?

*Measure specified to use data sources must be consistent with data sources entered in 2.19.*

Measure tested with data from

☐abstracted from paper record

☐administrative/management

☐claims

☐instrument-based

☐assessment

☐clinical database/registry

☐abstracted from electronic health records (EHRs)

☐electronic clinical quality measure (eCQM) Health Quality Measure Format (HQMF) implemented in EHRs/health information technology

☐other (specify) Click or tap here to enter text.

4.2 Identify the Specific Dataset

*If using an existing dataset, identify the dataset. The dataset used for testing must be consistent with the measure specifications for target population and measured entities (e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home Minimum Data Set [MDS] home health Outcome and Assessment Information Set [OASIS], clinical registry).*

4.3 Testing Data

4.3.1 What Are the Dates of the Data Used in Testing?

*Enter the date range for the testing data.*

4.3.2 What Levels of Analysis Did the Measure Developer Test?

*Provide testing for all levels specified and intended for measure implementation (e.g., individual clinician, hospital, health plan).*

Measure specified to measure performance of *(must be consistent with data sources entered in 2.19)*

☐individual clinician

☐group/practice

☐hospital/facility/agency

☐health plan

☐accountable care organization

☐geographic population

☐other (specify) Click or tap here to enter text.

Measure tested at level of

☐individual clinician

☐group/practice

☐hospital/facility/agency

☐health plan

☐accountable care organization

☐geographic population

☐other (specify) Click or tap here to enter text.

4.3.3 How Many and Which Measured Entities Were Included in the Testing and Analysis?

*Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if using a sample, describe the selection criteria for inclusion in the sample.*

4.3.4 How Many and Which Patients Were Included in the Testing and Analysis?

*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if using a sample, describe the selection criteria for patient inclusion in the sample. If there is a minimum case count used for testing, reflect the minimum in the specifications.*

4.3.5 Sample Differences, if applicable

*If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusion, risk adjustment), identify how the data or sample differ for each aspect of testing reported.*

4.3.6 What Were the Social and Functional Risk Factors That Were Available and Analyzed?

*Describe social and functional risk factors; for example, patient-reported data (e.g., income, education, language), proxy variables when the measured entity does not collect social and functional risk data from each patient (e.g., census tract), or patient community characteristics (e.g., percentage of vacant housing, crime rate), which do not have to be a proxy for patient-level data.*

*Test measures for all data sources and specified levels of analyses.*

*Information on scientific acceptability should be sufficient for CMS and interested parties to understand to what degree the testing results for the measure meet evaluation criteria for testing.*

4.4 [Reliability Testing](https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/scientific-acceptability/reliability) (for reference only)

*Reliability testing demonstrates measure data elements are repeatable, producing the same results a high percentage of the time when assessed in the same population in the same time period, and/or the measure score is precise.* *For instrument-based measures (including PRO-PMs) and composite measures, demonstrate reliability for the computed performance score.*

*Reliability testing applies to both the data elements and computed measure score. For composite measures, must demonstrate reliability of the computed performance score. Examples of reliability testing for data elements include inter-rater/abstractor or intra-rater/abstractor studies, internal consistency for multi-item scales, and test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).*

*If the measure developer empirically tested for accuracy/correctness (i.e., validity) of data elements, there is no requirement for separate reliability testing of data elements. In 4.4.1, check patient/encounter level; in 4.4.2, enter “refer to section 4.5 for validity testing of data elements”; and skip 4.4.3 and 4.4.4.*

4.4.1 Level of Reliability Testing

*At what level of reliability was testing conducted? (check all that apply)*

☐patient/encounter level (data element level) (e.g., inter-abstractor reliability)

☐accountable entity level (measure score level) (e.g., signal-to-noise analysis)

4.4.2 Method of Reliability Testing

*Describe the method of reliability testing for each level used, as identified in 4.4.1. Do not just name the method. What type of error is it testing? Provide the statistical analysis you used.*

4.4.3 Statistical Results from Reliability Testing

*What were the statistical results from reliability testing for each level, as identified in 4.4.1? Examples include percent agreement and kappa for the critical data elements, and distribution of reliability statistics from a signal-to-noise analysis. Provide reliability statistics and assessment of adequacy in the context of norms for the test conducted.*

4.4.4 Interpretation

*What is your interpretation of the results in terms of demonstrating reliability? What do the results mean and what are the norms for the test conducted?*

4.5 [Validity Testing](https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/scientific-acceptability/validity) (for reference only)

*Validity testing demonstrates the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality**. For instrument-based measures, including PRO-PMs and composite measures, demonstrate validity for the computed performance score.*

*Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to, testing hypotheses the measures scores indicate quality of care (e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method); correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate for measures not yet in a CMS program if accomplished by identified experts through a systematic and transparent process explicitly addressing whether performance scores resulting from the measure as specified and can distinguish good from poor quality.* *Provide/discuss the degree of consensus and any areas of disagreement.*

4.5.1 Level of Validity Testing

*At what level(s) of validity was testing conducted? (check all that apply)*

☐patient/encounter level (data element level)

☐accountable entity level (measure score level)

4.5.2 Type of validity testing

*Select the type or types of validity testing conducted.*

☐empirical validity testing

☐systematic assessment of face validity of quality measure score as an indicator of quality or resource use (i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance)

*Provide empirical validity testing at the time of maintenance review; if not possible, provide justification.*

4.5.3 Method of Validity Testing

*For each level tested, describe the method of validity testing and what it tests. Do not just name the method; please describe the steps and what the measure developer tested (e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected, statistical analysis used).*

4.5.4 Statistical Results from Validity Testing

*Provide statistical results and assessment of adequate validity (e.g., correlation, t test).*

4.5.5 Interpretation

*What is your interpretation of the results in terms of demonstrating validity? What do the results mean and what are the norms for the test conducted?*

4.6 [Exclusions Analysis](https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/scientific-acceptability/exclusions) (for reference only)

*Support exclusions by the clinical evidence and note sufficient frequency to warrant inclusion in the specifications of the measure. Examples of evidence an exclusion distorts measure results include frequency of occurrence, variability of exclusion across measured entities, and sensitivity analyses (with and without the exclusion). If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence the exclusion impacts performance on the measure; in such cases, specify the measure so the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). Patient preference is not a clinical exception to eligibility and measured entities interventions may influence patient preference.*

*If there are no exclusions, indicate this section is not applicable and skip to 4.7.*

4.6.1 Method of Testing Exclusions

*Describe the method of testing the exclusions and what it tests. Do not just name the method; describe the steps and what the measure developer tested (e.g., whether the exclusions affect overall performance scores); and statistical analysis used.*

4.6.2 Statistical Results from Testing Exclusions

*What were the statistical results from testing the exclusions? Include overall number and percentage of individuals excluded, frequency distribution of the exclusions across measured entities, and impact on quality measure scores.*

4.6.3 Interpretation

*What is your interpretation of the results in terms of demonstrating there is a need for exclusions to prevent unfair distortion of performance results (i.e., the value outweighs the burden of increased data collection and analysis)? If patient preference is an exclusion, specify the measure so the effect on the performance score is transparent (e.g., scores with and without the exclusion).*

4.7 Risk Adjustment or Stratification for Outcome or Resource Use Measures (for reference only)

*For outcome and other measures (e.g., resource use, cost), specify an evidence-based risk adjustment strategy (e.g., risk model, risk stratification), is based on patient factors (including clinical, sociodemographic, and functional factors) influencing the measured outcome, are present at start of care, are not associated with the quality of care, and have demonstrated adequate discrimination and calibration.*

*Do not specify risk factors influencing outcomes as exclusions. Measure developers should consider both stratification and risk adjustment of measures by risk factors (clinical, social, functional).*

*If this is not applicable, describe the rationale/data support for no risk adjustment/stratification.*

*If the measure is not an intermediate, or health outcome, PRO-PM, or resource use measure, state not applicable and skip to 4.8.*

*See also the* [*Risk Adjustment in Quality Measurement*](https://mmshub.cms.gov/sites/default/files/Risk-Adjustment-Quality-Measurement.pdf) *supplemental material.*

4.7.1 Method of Controlling for Differences

The method of controlling for differences in case mix is

☐no risk adjustment or stratification

☐statistical risk model with (specify number) risk factors

☐stratification by (specify number) risk categories

☐other (specify) Click or tap here to enter text.

*If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.*

4.7.2 Rationale for Why There Is No Need for Risk Adjustment

*If not risk-adjusting or stratifying an outcome or resource use measure, provide rationale and analyses to demonstrate there is no need for controlling for differences in patient characteristics (i.e., case mix) to achieve fair comparisons across measured entities.*

4.7.3 Conceptual, Clinical, and Statistical Methods/Model

*Describe the conceptual, clinical, and statistical methods and criteria used to select patient factors (i.e., clinical factors, social risk factors, or functional risk factors) used in the statistical risk model or for stratification by risk (e.g., potential factors identified in the literature and/or expert panel, regression analysis, statistical significance of p < 0.10, correlation of x or higher, patient factors should be present at the start of care and not related to disparities).* *Also, discuss any ordering of risk factor inclusion; for example, are social risk factors added after all clinical factors?*

4.7.4 Conceptual Model of Impact of Social and Functional Risks

*How was the conceptual model of the impact of social and functional risks of this outcome developed? Check all that apply.*

☐published literature

☐internal data analysis

☐other (specify) Click or tap here to enter text.

4.7.5 Statistical Results

*Describe the statistical results of the analyses used to select risk factors.*

4.7.6 Analyses and Interpretation in Selection of Social and Functional Risk Factors

*Describe the analyses and interpretation resulting in the decision to select social and functional risk factors (e.g., prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects). Also, describe the impact of adjusting for risk (or not) on measured entities at high or low extremes of risk.*

4.7.7 Method Used to Develop the Statistical Model or Stratification Approach

*Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach. Do not just name the method; describe the steps and identify the statistical analysis you used.*

*Provide the statistical results from testing the approach to controlling for differences in patient characteristics (i.e., case mix). If stratified, skip to 4.7.11.*

4.7.8 Statistical Risk Model Discrimination Statistics (e.g., c-statistic, R2)

4.7.9 Statistical Risk Model Calibration Statistics (e.g., Hosmer-Lemeshow statistic)

4.7.10 Statistical Risk Model Calibration—Risk decile plots or calibration curves

4.7.11 Results of Risk Stratification Analysis

4.7.12 Interpretation

*What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix), i.e., what do the results mean and what are the norms for the test conducted?*

4.7.13 Optional Additional Testing for Risk Adjustment

*While not required, this testing would provide additional support of adequacy of the risk model (e.g., testing of risk model in another data set, sensitivity analysis for missing data, other methods assessed).*

4.8 Identification of Meaningful Differences in Performance (for reference only)

*Data analysis of computed measure scores demonstrates methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance. With large enough sample sizes, small, statistically significant differences may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74% vs. 75%) is clinically meaningful, or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 vs. $5,025) is practically meaningful. Measures showing less-than-optimal performance may not demonstrate much variability across measured entities.*

*You may also describe the evidence of overall less-than-optimal performance. The intent of this section is to go beyond demonstrating a performance gap and address statistical significance, if possible.*

4.8.1 Method

*Describe the method for determining whether identification of statistically significant and clinically or practically meaningful differences in quality measure scores among the measured entities is possible. Do not just name the method, describe the steps and the statistical analysis you used. Do not just repeat the information provided related to performance gap in the section on importance, 3.2, Performance Gap.*

4.8.2 Statistical Results

*What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in quality measure scores across measured entities? For example, was an unexpected number and percentage of entities with scores significantly varying from the mean or some benchmark? How was meaningful difference defined?*

4.8.3 Interpretation

*What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? What do the results mean in terms of statistical and meaningful differences?*

4.9 Comparability of Multiple Data Sources/Methods (for reference only)

*If there is only one set of specifications, skip to 4.10.*

*If specifying multiple data sources/methods, there is demonstration they produce comparable results.*

*Measure developers should direct this item to risk-adjusted measures —with or without social or functional risk factors—or to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from patient record abstraction and a different set of specifications, e.g., claims or eCQMs). It does not apply to measures using more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and patient record abstraction for the numerator). There is no requirement for comparability when comparing performance scores with and without social/functional risk factors in the risk adjustment model. However, if there is no demonstration of comparability for measures with more than one set of specifications/instructions, submit the different specifications (e.g., for patient records vs. claims) as separate measures.*

4.9.1 Method

*Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources or specifications. Describe the steps―do not just name a method. Provide the statistical analysis used.*

4.9.2 Statistical Results

*What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications (e.g., correlation, rank order)?*

4.9.3 Interpretation

*What is your interpretation of the results in terms of demonstrating comparability of quality measure scores for the same entities across the different data sources or specifications? What do the results mean and what are the norms for the test conducted?*

4.10 Missing Data Analysis and Minimizing Bias (for reference only)

*Analyze and identify the extent and distribution of missing data (or nonresponse) and demonstrate there is no bias in performance results due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.*

4.10.1 Method

*Describe the testing method conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate there is no bias in performance results due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias. Describe the steps―do not just name a method. Provide the statistical analysis used, such as examples of evidence missing data distorts measure results include, but not limited to, frequency of occurrence and variability across measured entities.*

4.10.2 Missing Data Analysis

*What is the overall frequency of missing data, the distribution of missing data across measured entities, and the results from testing related to missing data (e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse)? If there is no empirical sensitivity analysis, identify the approaches considered for handling missing data and pros and cons of each.*

4.10.3 Interpretation

*What is your interpretation of the results in terms of demonstrating there is no bias in performance results due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias? What do the results mean in terms of supporting the selected approach for missing data, and what are the norms for the test conducted? If you did not conduct empirical analysis, provide the rationale for the selected approach for missing data.*

1. [**Feasibility**](https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/feasibility)

*This criterion assesses the extent to which the required data are readily available, retrievable without undue burden, and are implementable for performance measurement.*

5.1 Data Elements Generated as Byproduct of Care Processes

*How are the needed data elements generated to compute measure scores?*

Data used in the measure are (check all that apply)

☐generated or collected by and used by health care personnel during provision of care (e.g., blood pressure, laboratory value, diagnosis, depression score)

☐coded by someone other than the person obtaining original information (e.g., Diagnosis-Related Group, International Classification of Diseases, 10th Revision, Clinical Modification/Procedure Coding System codes on claims)

☐abstracted from a record by someone other than the person obtaining original information (e.g., chart abstraction for quality measure or registry)

☐other (specify) Click or tap here to enter text.

5.2 Electronic Sources

5.2.1 Data Elements Electronic Availability

*To what extent are the data elements needed for the measure available electronically* *(i.e., needed elements to compute quality measure scores are in defined, computer-readable fields)?*

☐All data elements are in defined fields in EHRs.

☐All data elements are in defined fields in electronic claims.

☐All data elements are in defined fields in electronic clinical data such as clinical registry, nursing home MDS, and home health OASIS.

☐All data elements are in defined fields in a combination of electronic sources.

☐Some data elements are in defined fields in electronic sources.

☐No data elements are in defined fields in electronic sources.

☐Data are patient/family reported information; may be electronic or paper.

5.2.2 Path to Electronic Capture

*If all data elements needed to compute the quality measure score are not from electronic sources, specify a credible, near-term path to electronic capture or provide a rationale for using other than electronic sources.*

5.2.3 eCQM Feasibility

*If not an eCQM, state N/A.*

*If this is an eCQM, provide a summary of the feasibility assessment in an attached file, e.g., the* [*eCQM Feasibility Scorecard*](https://p4qm.org/sites/default/files/2023-08/eCQM-Feasibility-Scorecard.xlsx)*, or make it available at a measure-specific URL.*

5.3 Data Collection Strategy

5.3.1 Data Collection Strategy Difficulties (optional)

*Describe difficulties as a result of testing or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, and other feasibility or implementation issues.*

*If the measure is instrument-based, consider the implications of burden for both individuals providing the data (e.g., patients, service recipients, respondents) and the measured entities.*

5.3.2 Fees, Licensing, Other Requirements

*Describe any fees, licensing, or other requirements to use any aspect of the measure as specified, such as the value or code set, the risk model, programming code, or algorithm. Please provide the fee schedule, if available. If none, state N/A.*

1. [**Usability and Use**](https://mmshub.cms.gov/measure-lifecycle/measure-testing/evaluation-criteria/usability)

*This criterion evaluates the extent to which intended audiences such as consumers, purchasers, measured entities, and policy makers can understand results of the measure and are likely to find them useful for decision-making. CMS expects use of CMS CBE-endorsed measures are in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to being in use for performance improvement.*

6.1 Usability

6.1.1 Improvement

*Refer to data provided in 3.2, Performance Gap, but do not repeat here. Discuss or document progress on improvement, such as trends in performance results; number and percentage of people receiving high-quality health care; and geographic area and number and percentage of accountable entities and patients included.*

*If there was no improvement demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale describing how to use the performance results to further the goal of high-quality, efficient health care for individuals or populations.*

6.1.2 Unexpected Findings

*Explain any unexpected findings—positive or negative—during implementation of this measure, including unintended impacts on patients.*

6.1.3 Unexpected Benefits

*Explain any unexpected benefits from implementation of this measure.*

6.2 Use

6.2.1 Current and Planned Use

*Select all uses that apply. Identify whether the use is current or planned.*

☐public reporting

☐public health or disease surveillance

☐payment program

☐regulatory and accreditation programs

☐professional certification or recognition program

☐quality improvement with external benchmarking to multiple organizations

☐quality improvement internal to a specific organization

☐not in use

☐use unknown

*For each current use listed, provide*

* *name of the program and sponsor*
* *URL for the program (if in current use)*
* *purpose*
* *geographic area*
* *number and percentage of accountable entities and patients included*
* *level of measurement*
* *setting*

6.2.1.1 Reasons for Not Publicly Reporting or Use in Other Accountability Application

*If not currently publicly reported or used in at least one other accountability application* *such as payment program, certification, or licensing, what are the reasons? Are there policies or actions of the measure developer and steward or accountable entities restricting access to performance results or impede implementation?*

6.2.1.2 Plan for Implementation

*If not currently publicly reported or used in at least one other accountability application, provide a credible plan for implementation within the expected time frames (i.e., any accountability application within 3 years and publicly reported within 6 years of initial endorsement). Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified time frames. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*

6.2.2 Feedback on the Measure by Measured Entities or Others

6.2.2.1 Technical Assistance Provided During Development or Implementation

*Describe the provision of performance results, data, and assistance with interpretation to measured entities or other users during development or implementation.*

*How many and which types of measured entities and/or others did you include? If you only included a sample of measured entities, describe the full population and describe the selection of the sample.*

6.2.2.2 Technical Assistance with Results

*Describe the process(es) involved, including when/how often you provided results, what data you provided, what educational/explanatory efforts you made, etc.*

6.2.2.3 Feedback on Measure Performance and Implementation

*Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how you obtained feedback.*

6.2.2.4 Feedback from Measured Entities

*Summarize the feedback obtained from measured entities.*

6.2.2.5 Feedback from Other Users

*Summarize the feedback obtained from other users.*

6.2.2.6 Consideration of Feedback

*Describe how you considered the feedback described in 6.2.2.3 when developing or revising the measure specifications or implementation, including whether you modified the measure and why or why not.*

***Additional Information***

***Appendix***

*Provide supplemental materials in an appendix.*

*Organize all supplemental materials, such as data collection instrument or methodology reports, in one file with a table of contents or bookmarks. Indicate if material pertains to a specific MIJF item number. Provide requested information in the MIJF and any measure testing attachment(s). There is no guarantee of review of supplemental materials. Indicate whether supplemental materials are available at a measure-specific web page (URL identified in 3.1 in the MIJF), available in attached file, or no supplemental materials.*

***Other Additional Information***

Ad.1. Working Group/Expert Panel Involved in Measure Development

*List the working group/panel members’ names and organizations.*

*Describe the members' role in measure development.*

***Measure Developer/Steward Updates and Ongoing Maintenance***

Ad.2. First Year of Measure Release

Ad.3. Month and Year of Most Recent Revision

Ad.4. What is your frequency for review/update of this measure?

Ad.5. When is your next scheduled review/update for this measure?

Ad.6. Copyright Statement

Ad.7. Disclaimers

Ad.8. Additional Information/Comments

*[Please delete this list of references and replace with your own references before submission]*

**References**

Agency for Healthcare Quality and Research. (n.d.). *Evidence-based practice centers*. Retrieved January 8, 2024, from <https://effectivehealthcare.ahrq.gov/about/epc>

Centers for Medicare & Medicaid Services. (n.d.). *Creating accessible products*. Retrieved January 8, 2024, from <https://www.cms.gov/es/node/1549751>

Grading of Recommendations Assessment, Development and Evaluation Working Group. (n.d.). *GRADE*. Retrieved January 8, 2024, from <http://www.gradeworkinggroup.org/>

Institute of Medicine. (2011). [*Finding what works in health care: Standards for systematic reviews*](https://doi.org/10.17226/13059)*.* The National Academies Press. <https://doi.org/10.17226/13059>

U.S. Preventive Services Task Force. (n.d.). *Grade definitions*. Retrieved January 8, 2024, from <https://www.uspreventiveservicestaskforce.org/Page/Name/grade-definitions>