National Quality Forum

Measure Testing Form for Trial Approval Program

Measure Title:
Date of Submission:
Type of Measure:

☐ Composite –	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

A measure submission that is to be considered for the Trial Approval Program must complete this form in its entirety. Either a test data set provided by the measure developer, or the use of the Bonnie tool is acceptable to provide preliminary testing results,

For <u>all</u> measures being submitted for potential acceptance into the Trial Approval Program, each section <u>must be filled out as completely as possible.</u>

Respond to <u>all</u> questions as instructed with answers immediately following the question. All information on testing of either a sample data set or results from Bonnie testing that can demonstrate, to the extent possible, the measure meets reliability and validity must be in this form

If you are unable to check a box, please highlight or shade the box for your response.

Maximum of 10 pages (*including questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***

Contact NQF staff regarding questions at trialmeasures@qualityforum.org

<u>DATA and SAMPLING INFORMATION</u>

1. DATA/SAMPLE USED FOR <u>PRELMINARY</u> TESTING OF THIS MEASURE

It is important that the measure developer use a data set to conduct preliminary testing in order to evaluate the measure logic and the inclusions/exclusions for the population used in the measure.

What type of data was used for testing? (*The measure developer must provide a test data set that will provide some initial information to be used for the evaluation, or the Bonnie testing tool can be used to create a sample data set using synthesized patients.)* Please indicate whether the test data set used was provided through the measure developer, or through the Bonnie tool.

If Bonnie was NOT used, please identify the specifications for the test dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured)*

What levels of analysis were tested (either through the test data set or Bonnie)? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*) *in order to determine its suitability for inclusion into the Trial Approval Program.*

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.4. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis)*

If the Bonnie testing tool was used to provide a sample data set, please refer to the guidance for Bonnie testing found at this link:
http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=80307
Bonnie testing results may be compiled into spreadsheet or table, which must be completed in its entirety, to the extent possible, in order to provide a basis for evaluation to determine the acceptability of the measure for inclusion in the Trial Approval program. Refer to this link for an example of formatting Bonnie results:
http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=81576

Any questions regarding the completion of this form can be directed to NQF Staff at trialmeasures@qualityforum.org.

RELIABILITY AND VALIDITY ASSESSMENTS

Note: The information provided in this next section is intended to aid the Standing Committee and other stakeholders in understanding to what degree the measure is both reliable and valid. While it is not possible to provide comprehensive results due to the lack of actual testing data, the developer needs to provide as much information as possible based on their interpretation of the results from the sample test data.

2.1 Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score. **What is your interpretation of the results in terms of demonstrating reliability**? (i.*e., what do the sample results mean and what are the norms for the test conducted?*) **Please summarize the plan for future testing of reliability if the measure is accepted into the Trial Approval Program.**

Include descriptions of:

Inter-abstractor reliability, and data element reliability of all critical data elements

Computation of the performance measure score (e.g., signal-to-noise analysis)?

2.2 Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score. **What is your interpretation of the results in terms of demonstrating validity**? (i.*e., what do the results mean and what are the norms for the test conducted?*). Please summarize the plan for future testing of validity if the measure is accepted into the Trial Approval Program. Include the method(s) of validity testing and what it will test (describe the steps—do not just name a method; what will be tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis will be used used)

2.3 Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion. **What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results?** (*i.e., the value outweighs the burden of increased data collection and analysis).* Please summarize the plan for future testing of exclusions if the measure is accepted into the Trial Approval Program. Describe the method of testing exclusions and what it will test (describe the steps—do not just name a method; what will be tested, e.g., whether exclusions affect overall performance scores; what statistical analysis will be used)

2.4 Risk Stratification (applicable ONLY to outcome or resource use measures). If an outcome or resource use measure will not be risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities. If risk adjustment/stratification is needed then please describe the conceptual/clinical and statistical methods and criteria that will be used to select patient factors (clinical factors or sociodemographic factors) that will be 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*)