**Measure Testing to Demonstrate Scientific Acceptability of Measure Properties**

**Measure Title**: Click here to enter measure title

**Date of Submission**: Click here to enter a date

**Type of Measure:**

|  |  |
| --- | --- |
| [ ]  Composite | [ ] Outcome |
| [ ] Cost/resource | [ ] Process |
| [ ] Efficiency | [ ] Structure |

|  |
| --- |
| This Word document template must be used to submit information for measure testing.* **For all measures, sections 1, 2a2, 2b2, 2b3, 2b5 must be completed**
* **For outcome or resource use measures**, section **2b4** also must be completed
* If specified for **multiple data sources** (e.g., claims and medical records), section **2b6** also must be completed
* Respond to all questions with answers immediately following the question (*unless meet the skip criteria or those that are indicated as optional*).
* Maximum of 10 pages (*incuding questions/instructions; do not change margins or font size; contact project staff if need more pages*)
* All information on testing to demonstrate meeting the [criteria for scientific acceptability of measure properties (2a,2b)](http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=66289) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
 |

**1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE**

*Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 7.*

**1.1. What type of data was used for testing**? (*Check all the sources of data identified in the measure specifications and data used for testing the measure*. *Testing must be provided for all the types of data specified and intended for measure implementation*)

|  |  |
| --- | --- |
| **Measure Specified to Use Data From:** | **Measure Tested with Data From:** |
| [ ] abstracted from paper record | [ ] abstracted from paper record |
| [ ] administrative claims | [ ] administrative claims |
| [ ] clinical database/registry | [ ] clinical database/registry |
| [ ] abstracted from electronic health record | [ ] abstracted from electronic health record |
| [ ] eMeasure implemented in electronic health record | [ ] eMeasure implemented in electronic health record |
| [ ] other: Click here to describe | [ ] other: Click here to describe |

**1.2. If used an existing dataset, identify the specific dataset** (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

**1.3. What are the dates of the data used in testing**? Click here to enter date range

**1.4. What levels of analysis** **were tested**? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)
[ ] **individual clinician** [ ] **group/practice** [ ] **hospital/facility/agency** [ ] **health plan**[ ] **other**: Click here to describe

**1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)**? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

**1.6. 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 a sample was used, describe how patients were selected for inclusion in the sample*)

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

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2a2. RELIABILITY TESTING**

***Note****: If accuracy/correctness (validity) of data elements was empirically tested*, *separate reliability testing of data elements is not required – report validity of data elements in 2b2*

**2a2.1. What level of reliability testing was conducted**? (*may be one or both levels*)
[ ]  **Critical data elements used in the measure** (*e.g., inter-abstractor reliability*)
[ ]  **Performance measure score** (e.g., *signal-to-noise*)

**2a2.2. For each level checked above, describe the method of reliability testing and what it tests** (*describe the steps―do not just name a method; what type of error does it test; what statistical analysis was used*)

**2a2.3. For each level checked above, what were the statistical results from reliability testing**? (e*.g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis and association with case volume*)

**2a2.4 What is your interpretation of the results in terms of demonstrating reliability**? (i*.e., what do the results mean and what are the norms for the test conducted?*)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2b2. VALIDITY TESTING**

**2b2.1. What level of validity testing was conducted**? (*may be one or both levels*)
[ ]  **Critical data elements**

[ ]  **Performance measure score**

[ ]  **Empirical validity testing**[ ]  **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance quality or resource use and can distinguish performance*)

**2b2.2. For each level checked above, describe the method of validity testing and what it tests** (*describe the steps―do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

**2b2.3. What were the statistical results from validity testing**? (*e.g., correlation; t-test, ANOVA*)

**2b2.4. 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?*)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2b3. EXCLUSIONS ANALYSIS**

**NA** [ ]  **no exclusions — *skip to #2b5***

**2b3.1. Describe the method of testing exclusions and what it tests** (*describe the steps―do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

**2b3.2. What were the statistical results from testing exclusions**? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

**2b3.3. 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.*  *Note:* ***If patient preference is an exclusion****, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE**

**2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified** (*describe the steps―do not just name a method; what statistical analysis was used)*

**2b5.2. What were the statistical results from testing the ability to identify differences in performance measure scores across measured entities?** (*at a minimum, the distribution of performance measure scores for the measured entities by decile/quartile, mean, std dev; preferably also number and percentage statistically different from mean or some benchmark, different form expected, etc*.)

**2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and clinically/practically meaningful differences in performance across measured entities?** (i*.e., what do the results mean and what are the norms for the test conducted?*)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***If not an intermediate or health outcome or resource use measure, this section can be deleted***

**2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES**

**2b4.1. What method of controlling for differences in case mix is used?**

[ ]  **Statistical risk model with** Click here to enter number of factors **risk factors**

[ ]  **Stratification by** Click here to enter number of categories **risk categories**

[ ]  **No risk adjustment or stratification**

[ ]  **Other,** Click here to enter description

**2b4.2. If an outcome or resource use measure is not 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**.

**2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select factors used in the statistical risk model or for stratification by risk** (*e.g., potential factors identified in literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher*)

**2b4.4. What were the statistical results of the analyses used to select risk factors?**

**2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach** (*describe the steps―do not just name a method; what statistical analysis was used*)

*Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below*.
***if stratified, skip to 2b4.9***

**2b4.6. Statistical Risk Model Discrimination Statistics:**

**2b4.7. Statistical Risk Model Calibration Statistics**:

**2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves**:

**2b4.9. Results of Risk Stratification Analysis**:

**2b4.10. 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*)

\***2b4.11.** **Optional Additional Testing** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods*)

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

***If only one set of specifications, this section can be deleted***

**2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS**

**Note***: This criterion is directed 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 medical records and a different set of specifications for claims). It does not apply to measures that use more than one type of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).*

**2b6.1. Describe the method of testing conducted to demonstrate equivalence of performance scores for the same entities across the different specifications** (*describe the steps―do not just name a method; what statistical analysis was used*)

**2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different specifications?** (*e.g., correlation, rank order*)

**2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different specifications?** (i*.e., what do the results mean and what are the norms for the test conducted*)