**National Quality Forum—Measure Testing (subcriteria 2a2, 2b2-2b7)**

**Measure Number** (*if previously endorsed*)**:** Click here to enter NQF number

**Measure Title**:

**Date of Submission**: 1/20/2014

**Type of Measure:**

|  |  |
| --- | --- |
| Composite – ***STOP – use composite testing form*** | Outcome (*including PRO-PM*) |
| Cost/resource | Process |
| Efficiency | Structure |

|  |
| --- |
| **Instructions**   * Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form. * **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.** * **For outcome and resource use measures**, section **2b4** also must be completed. * If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed. * Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed. * If you are unable to check a box, please highlight or shade the box for your response. * Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.*** * Contact NQF staff regarding questions. Check for resources at [Submitting Standards webpage](http://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx). |

|  |
| --- |
| **Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF’s evaluation criteria for testing.**  **2a2.** **Reliability testing** [**10**](#Note10) 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.  **2b2.** **Validity testing** [**11**](#Note11) 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.    **2b3.** 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; [**12**](#Note12)  **AND**  If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [**13**](#Note13)  **2b4.** **For outcome measures and other measures when indicated** (e.g., resource use):   * **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [**14**](#Note14)**,**[**15**](#Note15) and has demonstrated adequate discrimination and calibration   **OR**   * rationale/data support no risk adjustment/ stratification.   **2b5.** Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [**16**](#Note16) **differences in performance**;  **OR**  there is evidence of overall less-than-optimal performance.  **Higher scores indicate higher quality. Improvement opportunities exist in all areas.**  **2b6.** **If multiple data sources/methods are specified, there is demonstration they produce comparable results**.  **2b7.** For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data),analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.  **Notes**  **10.** Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).  **11.** 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 that 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 if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.  **12.** Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.  **13.** Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.  **14.** Risk factors that influence outcomes should not be specified as exclusions.  **15.** Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.  **16.** With large enough sample sizes, small differences that are statistically significant 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 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers. |

**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 1.7.*

**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 sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

|  |  |
| --- | --- |
| **Measure Specified to Use Data From:**  **(*must be consistent with data sources entered in S.23*)** | **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 (HQMF) implemented in EHRs | eMeasure (HQMF) implemented in EHRs |
| other: Click here to describe | other: Click here to describe |

**1.2. If an existing dataset was used, 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

2006-2010, 2012-2013

**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*)

|  |  |
| --- | --- |
| **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 |

**1.5. 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*)   
Approximately 75 hospitals were included in the initial three rounds of testing. An additional 73 hospitals were included in the 2012-2013 testing.

All hospitals are rural hospitals. Most are Critical Access Hospitals. They are located in 9 states across the US.

**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*)

Approximately 1500 patients were included in the 2006-2010 testing. Data details for the 2012-2013 is not yet available.

**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 – in 2a2.1 check critical data elements; in 2a2.2 enter “see section 2b2 for validity testing of data elements”; and skip 2a2.3 and 2a2.4.*

**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; data element reliability must address ALL critical data elements*)  
 **Performance measure score** (e.g., *signal-to-noise analysis*)  
  
**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*)

To assure that data for the measures were being abstracted consistently across all hospitals and across all abstractors, inter-rater reliability (IRR) testing was included in three field tests 2006, 2008, 2010. IRR testing ensures data quality and abstraction consistency by re-auditing a sample of abstracted medical records from each hospital. An element-to-element comparison is performed to help identify problem areas in the measure definitions or the measure applications. This also gives the hospitals an opportunity to make comments about measures that may be confusing and need more explanation.

The abstraction services coordinator for Stratis Health conducted inter-rater reliability testing for the first field test, staff from the Washington Rural Health Quality Network conducted inter-rater reliability testing for the second field test and Stroudwater and University of Minnesota conducted the IRR for the third field test. A sample of 1-3 records from each hospital was re-audited and findings were compared with those of each hospital auditor. Each auditor was contacted by phone to discuss the differences in findings and clarify definitions.

All hospitals that completed the inter-rater reliability process did well on the abstraction. Those individuals without previous experience in data collection from medical records required additional assistance with the collection of demographic data. The Emergency Department Transfer Communication Measures were new to all participants and required additional clarification for most auditors. The two main areas of difficulty in data abstraction were collecting patient arrival/discharge times and the recording of when vital signs were collected. Clarification of the abstraction definitions and applications were provided to the individual auditors in a phone discussion with examples from their own records.

**2a2.3. For each level of testing 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*)

For the first quarter of data for field test one, 70 transfer records were abstracted at 23 hospitals in three states. For 68% of those records, the hospital abstractors’ findings agreed 100% with the QIO staff abstraction. In the second quarter of field test one, another 60 transfer records charts were abstracted at 19 hospitals. For 82.4% of those records, the hospitals’ abstraction findings agreed 100% with the QIO staff abstraction. The number of inconsistencies in abstraction decreased by more than 50% from the first quarter to the second quarter. For field test two, on-site IRR was conducted shortly after the training. Sixty transfer records were abstracted and nearly all elements of all records matched the trainer’s abstractions. Clarification on admission dates and times was required.

**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?*)  
Reliability results were interpreted to mean that initial understanding of measurement elements was high. Little review or reinforcement was needed for trainees. Little revision or clarification of materials was indicated.

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

**2b2. VALIDITY TESTING**

**2b2.1. What level of validity testing was conducted**? (*may be one or both levels*)  
 **Critical data elements** (*data element validity must address ALL 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 on quality or resource use and can distinguish good from poor performance*)

**2b2.2. For each level of testing 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)*  
The following data supports the face validity of the measures.

The measures were developed based on a review of the quality measurement literature and consultation with experts in the field. Existing quality indicator and performance measurement systems (e.g., those developed by the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO), Agency for Healthcare Research and Quality, National Quality Forum, Centers for Medicare and Medicaid Services and four rural-oriented performance measurement systems), were reviewed with attention to identifying high priority areas for rural hospitals (e.g., emergency room stabilization and transfer) that were not currently being systematically collected.

A technical expert panel comprised of national experts in rural health and quality measurement was convened to review the findings of field test 1 and provide input about the inclusion of specific measures in a revised set of quality measures relevant for rural hospitals. It was the consensus of our research team to build the expert panel from a portion of the national expert panel that made the original measurement recommendations. The original panel included representatives from the rural physician, hospital, nursing, and pharmacy communities, and from the Joint Commission, NQF, Leapfrog, the Agency for Healthcare Research and Quality, Office of Rural Health Policy, QIOs, and purchaser coalitions. Eight members of the original panel were invited to participate. The additional members of the expanded panel included hospital participants from the current study, emergency room experts, a consultant who worked closely with the development of the Continuity of Care record (CCR), an expert in rural health policy, representatives from the American Academy of Family Physicians and the American Hospital Association, and an expert in medication safety and medication errors.

The 16 members of the expert panel were asked to rate the newly developed measures on three criteria using a 5-point Likert scale with five representing very useful or very prevalent. The three criteria were the prevalence in small rural hospitals, the internal usefulness for small rural hospitals and the external usefulness for small rural hospitals. The measures that were rated included the emergency department transfer communication measures. Fifteen of the sixteen panel members rated the measures.

To select measures relevant for quality improvement within a rural hospital, we identified measures that the expert panel rated higher than 4 on the five-point scale for internal or external usefulness and higher than 3 on the five-point scale for prevalence. All measures met the prevalence standard. All of the Emergency Department transfer communication measures were rated higher than 4 for internal and external usefulness.

To address the validity of the measure: Please see this 2013 Journal of Rural Health Article which describes the most recent expert panel review of these measures.

[J Rural Health.](http://www.ncbi.nlm.nih.gov/pubmed/23551646) 2013 Spring;29(2):159-71. doi: 10.1111/j.1748-0361.2012.00420.x. Epub 2012 Aug 1.

**Rural relevant quality measures for critical access hospitals.**

[Casey MM](http://www.ncbi.nlm.nih.gov/pubmed?term=Casey%20MM%5BAuthor%5D&cauthor=true&cauthor_uid=23551646)1, [Moscovice I](http://www.ncbi.nlm.nih.gov/pubmed?term=Moscovice%20I%5BAuthor%5D&cauthor=true&cauthor_uid=23551646), [Klingner J](http://www.ncbi.nlm.nih.gov/pubmed?term=Klingner%20J%5BAuthor%5D&cauthor=true&cauthor_uid=23551646), [Prasad S](http://www.ncbi.nlm.nih.gov/pubmed?term=Prasad%20S%5BAuthor%5D&cauthor=true&cauthor_uid=23551646).

<http://www.ncbi.nlm.nih.gov/pubmed/23551646>

The measures have continued to be used for process improvement in multiple states.

1. The following paragraphs describe three of the current projects which use these measures:
   1. As part of the Minnesota Statewide Quality Reporting and Measurement System established by state health care reform legislation in 2008,[[1]](#endnote-1) the Critical Access Hospitals (CAHs) in Minnesota were required to report data on the ED patient transfer communication measures to the Minnesota Department of Health starting in February 2012.10 This study examines the implementation of the ED patient transfer communications measures in the 79 CAHs in Minnesota in 2011-2013. Attached is an abstract of a submitted article describing the project.
   2. The measures are currently being used in 8 states as a pilot prior to inclusion in the Phase 3 of the Medicare Beneficiary Quality Improvement project (MBQIP).

Project Overview: The Critical Access Hospital (CAH) Emergency Department Transfer

Communication QIO special project is intended to build QIO capacity to support and improve the care delivered by CAHs focusing on transfer communication from the CAH emergency department.

This 12-month QIO special innovation project will provide and pilot test QIO resources for supporting CAHs to be trained to collect and report the ED Transfer Communication measures, identify gaps and opportunities for improvement, and begin planning for improving the transfer communication process and results. Eight QIOs are participating, supporting CAH participation in the following states: Iowa, Missouri, Nebraska, Maine, Oklahoma, West Virginia, Wisconsin, Wyoming.  Stratis Health, the Minnesota QIO, is coordinating and providing tools, training, and resources for each QIO to utilize in working with the participating CAHs in their respective states.  The project timeline is 8/1/13  - 7/31/14.  CAHs will begin collecting data in late 2013.

* 1. The project aligns with the HRSA funded federal Office of Rural Health Policy's (ORHP) Medicare Rural Hospital Flexibility Program (Flex) priorities in the Medicare Beneficiary Quality Improvement Project (MBQIP).  The ED Transfer Communication measure will be included in phase three of MBQIP, which launches in 2014.   QIOs are expected to work collaboratively with their state Flex offices.

**2b2.3. What were the statistical results from validity testing**? (*e.g., correlation; t-test*)  
none

**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?*)  
Face validity was upheld with repeated review by expert panels and hospital based experts.

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

**2b3. EXCLUSIONS ANALYSIS**

**NA**  **no exclusions — *skip to section*** [***2b4***](#section2b4)

**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*)

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

**2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES**  
***If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section*** [***2b5***](#section2b5)***.***

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

**No risk adjustment or stratification**

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

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

**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 patient 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*)

**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***](#question2b49)

**2b4.6. Statistical Risk Model Discrimination Statistics** (*e.g., c-statistic, R-squared*)**:**

**2b4.7. Statistical Risk Model Calibration Statistics** (*e.g., Hosmer-Lemeshow statistic*):

**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 for Risk Adjustment** (*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 that were assessed*)

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

**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? Do not just repeat the information provided related to performance gap in 1b)*

Statistical differences can be determined using one way anova when sample sizes allow. Small rural hospitals experience large variations in observations resulting in few significant findings.

**2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities?** (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)  
Few significant differences exist with small sample sizes. Differences were observed at state levels.

**2b5.3. 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 across measured entities?** (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

Because communication policy and management does not occur at the state level the these findings do not influence outcomes or improvement efforts.

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

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

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

**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 record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

**2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/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 data sources/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 data sources/specifications?** (i*.e., what do the results mean and what are the norms for the test conducted*)

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

**2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS**

**2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased** due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps―do not just name a method; what statistical analysis was used*)

**2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, 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 no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

**2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased** due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i*.e., 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 no empirical analysis, provide rationale for the selected approach for missing data*)

1. [↑](#endnote-ref-1)