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

**Measure Number** (*if previously endorsed*)**:** 2079

**Measure Title**: HIV medical visit frequency

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

**Type of Measure:**

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

**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:** |
| [x]  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*).

We utilized the multisite HIV Research Network (HIVRN), a consortium of community and academic HIV providers care sites, linked by a centralized Data Coordinating Center (DCC). The HIVRN has 11 participating treatment sites (10 adolescent/adult sites and 1 pediatric site). The sites are representative of both academic and community-based HIV care; of the 4 major geographic divisions of the U.S. of the demographic diversity of HIV infection across the U.S. and of the insurance status and coverage types typical of the population in care. The measurement periods included calendar years 2007-2008, 2008-2009, 2009-2010, and 2014-2015. More information can be found on the HIVRN website regarding site locations, additional data, and more.

All of the patients in the HIVRN dataset have a diagnosis of HIV. Patients were included, regardless of age, in each measurement period if they had a medical visit in the first 6 months of the measurement period and did not die during the measurement period. The following lists the number of patients included for each measurement period. Due to resource constraints, 2011-2013 were not included in the analysis to allow for inclusion of the most recent measurement period for this measure (2014-2016) with limited analysis available.

**1.3. What are the dates of the data used in testing**? 2010-2014

**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 |
| [x]  hospital/facility/agency | [x]  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*)

We utilized the multisite HIV Research Network (HIVRN), a consortium of community and academic HIV providers care sites, linked by a centralized Data Coordinating Center (DCC). The HIVRN has 11 participating treatment sites (10 adolescent/adult sites and 1 pediatric site). The sites are representative of both academic and community-based HIV care; of the 4 major geographic divisions of the U.S. of the demographic diversity of HIV infection across the U.S. and of the insurance status and coverage types typical of the population in care. The measurement periods included calendar years 2007-2008, 2008-2009, 2009-2010, and 2014-2015. More information can be found on the HIVRN website regarding a site location, additional data, and more.

All of the patients in the HIVRN dataset have a diagnosis of HIV. Patients were included, regardless of age, in each measurement period if they had a medical visit in the first 6 months of the measurement period and did not die during the measurement period. The following lists the number of patients included for each measurement period. Due to resource constraints, 2011-2013 were not included in the analysis to allow for inclusion of the most recent measurement period for this measure (2014-2016) with limited analysis available.

**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*)

The data for measure testing were collected via the Ryan White HIV/AIDS Program Services Report (RSR), which is HRSA HIV/AIDS Bureau's primary source of annual, client-level data collected from more than 2,000 funded grant recipients and subrecipients. The RSR is inclusive of the overall RWHAP client population and key priority populations served by RWHAP. The average number of patients per provider each year ranged from 384 to 411, shown in the table below. Descriptive characteristics (e.g., age, race/ethnicity, gender) for the patient population are shown in the subsequent table by year.

|  |  |
| --- | --- |
| Year | Number of patients included |
| 2007-2008 | 15,790 |
| 2008-2009 | 16,881 |
| 2009-2010 | 17,687 |
| 2014-2015 | 15,049 |

Provider-level medical visit frequency performance scores, 2014-2015

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Provider Site | Total N | Percent of patients with a medical visit in each six month segment of the measurement period | Lower confidence interval | Upper confidence interval |
| A  | 399 | 55.13 | 50.22 | 59.95 |
| B | 1910 | 63.24 | 61.05 | 65.38 |
| C | 1425 | 68.21 | 65.74 | 70.57 |
| D | 1490 | 68.45 | 66.05 | 70.76 |
| E | 1276 | 68.8 | 66.21 | 71.92 |
| F | 4549 | 70.93 | 69.6 | 72.24 |
| G | 630 | 78.88 | 75.52 | 81.37 |
| H | 745 | 79.19 | 76.12 | 81.89 |
| I | 1582 | 79.45 | 77.39 | 81.95 |
| J | 452 | 82.74 | 78.97 | 85.95 |
| K | 591 | 83.76 | 80.55 | 86.51 |

Summary statistics for proportion of 2014-2015 patients meeting the numerator criteria across providers.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 2007-2008 | 2008-2009 | 2009-2010 | 2014-2015 |
| Minimum | 47.1% | 42.5% | 50.1% | 55.1% |
| Maximum | 86.1% | 83.1% | 82.8% | 83.8% |
| Mean | 66.7% | 67.73% | 68.9% | 72.6% |
| 25th percentile | 59.7% | 59.9% | 63.4% | 68.2% |
| 50th percentile | 70.6% | 66.2% | 67.7% | 70.9% |
| 75th percentile | 78.2% | 75.5% | 74.6% | 79.5% |

**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**.

HIV Research Network (HIVRN) was the sole source of data for the testing.

**1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).**

The patient-level sociodemographic variables included in the analysis include the following: Age, race/ethnicity; gender; transmission risk; and health care coverage.

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

**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*)
[x]  **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**

Reliability was calculated according to the methods outlined in a technical report prepared by J.L. Adams for the National Committee for Quality Assurance titled “The Reliability of Provider Profiling: A Tutorial” (RAND Corporation, TR-653-NCQA, 2009). In this context, reliability represents the ability of a measure to confidently distinguish the performance of one physician from another. As discussed in the report: “Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of variability in measured performance that can be explained by real differences in performance. There are 3 main drivers of reliability; sample size, differences between physicians, and measurement error.”

According to this approach, reliability is estimated with a beta-binomial model. The beta-binomial model is appropriate for measuring the reliability of pass/fail measures such as those proposed here. Reliability scores vary from 0.0 to 1.0, with a score of zero indicating that all variation is attributable to measurement error (noise, or individual accountable entity variance) whereas a reliability of 1.0 implies that all variation is caused by real difference in performance across accountable entities.

As discussed in the technical report, there is not a clear cut-off for minimum reliability level. Values above 0.7, however, are considered sufficient to see differences between some physicians (or clinics) and the mean, and values above 0.9 are considered sufficient to see differences between pairs of physicians (in this case clinics).

Clinic-specific reliability results for the “Medical visit frequency” measure are detailed in the Table below. Clinic-specific reliability is consistently greater than 0.9, and thus can be considered to be very good. Clinic-specific reliability was also calculated for 2008 and 2009. Results were consistent with results from 2010 and are not shown here.

**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*)

**Table 1: Clinic-Specific Reliability for Medical Visit Frequency Measure** – Year 2010

Between-clinic variance: 0.0072

Clinic n percent Reliability

A 2605 76.0 0.99

B 719 78.2 0.97

C 746 68.0 0.96

D 1888 74.1 0.99

E 327 52.3 0.90

F 1320 65.2 0.98

G 436 64.0 0.93

H 1217 50.1 0.97

I 1436 69.6 0.98

J 1742 66.5 0.98

K 444 61.5 0.93

L 3177 67.4 0.99

M 1102 73.8 0.98

Pediatric 528 82.8 0.96

**Median 0.97 (Range 0.90-0.99)**

**2b. VALIDITY. Validity, Testing, including all Threats to2a2.4 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?*)

Clinic-specific reliability results for the “Medical visit frequency” measure are detailed in the table above. Clinic-specific reliability is consistently greater than 0.9, and thus can be considered to be very good. Clinic-specific reliability was also calculated for 2008 and 2009. Results were consistent with results from 2010 and are not shown here.

**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**[x]  **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)*

1. Face validity for the measure was established through a technical work group empaneled for the development of the measure. The technical work group consisted of leading researchers and providers in HIV care and treatment as well as governmental and nongovernmental public health officials from across the country. The technical work group used a modified Delphi process whereby experts presented the most current research to the work group members. The work group members discussed each of the presentations and identified data elements for each measure. The work group members voted on the domains for the proposed measures. The vote was based on importance, ability to assess quality care, feasibility to implement measure, and use in quality improvement activities (e.g. ability to improve measure score). The votes were tallied and draft components of the measures (including data elements) were returned to the work group for additional voting via survey. Consensus was reach when a simple majority agreed on the final set of measures.

Technical work group members:

Bruce Agins, NYS DOH AIDS Institute, New York, NY

Judy Bradford, Fenway Community Health, Boston, MA

John Brooks, CDC, Atlanta, GA

Karen Brudney, Columbia University, New York, NY

Laura Cheever, HEALTH RESOURCES AND SERVICES ADMINISTRATION HAB, Rockville, MD

Nikki Cockern, Wayne State University, Detroit, MI

Chinazo Cunningham, Montefiore Medical Center, New York, NY

William Cunningham, UCLA, Los Angeles, CA

Julie Dombrowski, University of Washington, Seattle, WA

Edward Gardner, Denver Health, Denver, CO

Elvin Geng, UCSF, San Francisco, CA

Thomas Giordano, Baylor College of Medicine, Houston, TX

Barb Gripshover, Cleveland ACT UP, Cleveland, OH

Deborah Konkle Parker, University of Mississippi, Jackson, MS

Tim Long, Alliance Chicago, Chicago, IL

Cheryl Lynn-Besch, Louisiana State University, New Orleans, LA

Julio Marrero, COSSMA, San Juan, PR

Brian Montague, Brown University, Providence, RI

Michael Mugavero, University of Alabama, Birmingham, AL

Sylvia Naar King, Wayne State University, Detroit, MI

Josiah Rich, Brown University, Providence, RI

Allan Rodriguez, Miami University, Miami, FL

Amy Sitapati, UCSD, San Diego, CA

Avnish Tripathi, University of South Carolina, Charleston, SC

Gregory Winstead, Christian Community Health Center, Chicago, IL

1. Face validity of the performance score was gained through a structured presentation (two identical presentations) to a national audience of Ryan White HIV/AIDS Program grant recipients, subrecipients, and stakeholders. Health Resources and Services Administration presented detailed information (e.g. work group process, numerator, denominator, exclusions, and data elements). The national audience includes organization that would use the measure on a routine basis for assessing quality of care and quality improvement purposes; providers of HIV health care; measurement experts and researchers; and people living with HIV. Four hundred and forty-five individuals participated in the webinars. Ryan White HIV/AIDS Program grant recipients, subrecipients, and stakeholders were invited to provide feedback about the implement the measure within their clinical quality management program including ability of the measure to assess quality care and feasibility of implementing the measure. Written feedback was submitted and reviewed.

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

1. The technical work group developed a measure that could be implemented to assess and improvement quality of care by Ryan White HIV/AIDS Program grant recipients and subrecipients.
2. Sixty-nine individuals/organizations submitted 239 pieces of comments. Seventeen comments were received regarding this measure. The comments included continuing efforts to align this measure across federal programs; availability of benchmarking data; clarification on measure details; and use in special populations (e.g. youth and young adults). Heath Resources and Services Administration did not receive any comments encouraging the discontinuation of the measure, inability of measure to assess quality of care; or inability to implement the measure.

**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?*

1. The technical work group was represented of the Ryan White HIV/AIDS Program grant recipients, subrecipients, and stakeholders and included clinical providers, researchers, and clinical quality management staff. The technical work group agreed upon a measure that could assess and improvement the quality of HIV care.
2. Health Resources and Services Administration provided detailed information about this measure to a large portion of the Ryan White HIV/AIDS Program grant recipients, subrecipients, and national partners (445 participants). Many comments (239) were received as a result of the presentations, which indicated a high degree of engagement with Health Resource and Services Administration regarding performance measures. Nearly 10% of the comments (17) were directly in response to this measure. None of the comments indicated that the measure should be discontinued, could not assess quality of care, or could not be implemented. No changes to the measure were made based on the feedback receive. Frequently asked questions were developed based on the feedback (available at [http://hab.Health Resources and Services Administration .gov/clinical-quality-management/performance-measure-portfolio](http://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio)).

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

**2b3. EXCLUSIONS ANALYSIS (FOR MEASURS WITH EXCLUSIONS --- gap in visits and medical visit frequency)**

**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*)

This measure has one exclusions – patient death during the measurement period. Due to constraints, we were not able to test the impact of the exclusion on this measure. It is important to note that patient mortality has reduced dramatically over the years primarily in relation to the development and dissemination of HIV antiretroviral therapy. Thus, we do not anticipate a significant number of patients that would be excluded from the measure.

Based on data from other measures, less than 1% of patients were excluded due to death each year.

**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*)

Due to constraints, we were not able to test the impact of the exclusion on this measure.

**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*)

Due to constraints, we were not able to test the impact of the exclusion on this measure.

**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?**

[x]  **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

**2b4.1.1 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.** N/A

**2b4.2. If an outcome or resource use component 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**.

N/A

**2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

N/A

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

**2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)** N/A

**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*) N/A
*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*)**:** N/A

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

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

**2b4.9. Results of Risk Stratification Analysis**: N/A

**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*) N/A

**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*) N/A

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

**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)*

To examine meaningful differences in performance, we examined the distribution of the proportion of patients with achieving medical visit frequency across providers, by year. Performance scores were broken into the percentiles to better characterize the gaps that remain across providers. Moreover, performance scores were examined with respect to National HIV/AIDS Strategy 2020 Indicator 5: Increase the percentage of persons with diagnosed HIV infection who are retained in

HIV medical care to at least 90 percent. (The National HIV/AIDS Strategy 2020 retention indicator definition is different, yet provides a benchmark.)

**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*)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 2007-2008 | 2008-2009 | 2009-2010 | 2014-2015 |
| Minimum | 47.1% | 42.5% | 50.1% | 55.1% |
| Maximum | 86.1% | 83.1% | 82.8% | 83.8% |
| Mean | 66.7% | 67.73% | 68.9% | 72.6% |
| 25th percentile | 59.7% | 59.9% | 63.4% | 68.2% |
| 50th percentile | 70.6% | 66.2% | 67.7% | 70.9% |
| 75th percentile | 78.2% | 75.5% | 74.6% | 79.5% |

**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?*)

The table above demonstrates meaningful variability across providers, allowing for the identification of meaningful differences across sites. Specifically, the measure is able to detect providers with better or worse than median performance scores. Focusing on the 2014-2015 data, the 25th percentile is 68.2% and the 75th percentile is 79.5%, which is more than 10 points higher than the 25th percentile. Further there is an even greater spread between the minimum and maximum percentages. While the gap appears to be narrowing over time, a meaningful difference of remains, demonstrating the value of the measure in identifying sites based on poor performance relative to the top performers.

**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 item is directed to measures that are risk-adjusted (with or without SDS 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 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).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

**2b6.1. Describe the method of testing conducted to compare 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*) N/A

**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*) N/A

**2b6.3. What is your interpretation of the results in terms of the differences in 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*) N/A
**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**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*)

Due to constraints, we did not analyze missing data.

**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*)

Because the data used in this measure are routinely collected and stored in health records as well as used for billing, we do not feel there is a significant amount of missing data or even enough to bias the results.

**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*)

N/A