MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-Meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2079
Measure Title: HIV medical visit frequency
Measure Steward: Health Resources and Services Administration - HIV/AIDS Bureau
Brief Description of Measure: Percentage of patients, regardless of age, with a diagnosis of HIV who had at least one medical visit in each 6-month period of the 24-month measurement period with a minimum of 60 days between medical visits. A medical visit is any visit in an outpatient/ambulatory care setting with a nurse practitioner, physician, and/or a physician assistant who provides comprehensive HIV care.
Developer Rationale: Poor retention in care during the first year of outpatient medical care is associated with delayed or failed receipt of antiretroviral therapy, delayed time to virologic suppression and greater cumulative HIV burden, increased sexual risk transmission behaviors, increased risk of long-term adverse clinical events, and low adherence to antiretroviral therapy. Early retention in HIV care has been found to be associated with time to viral load suppression and 2-year cumulative viral load burden among patients newly initiating HIV medical care (8). In this study, each "no show" clinic visit conveyed a 17% increased risk of delayed viral load suppression. A dose- response relationship has been shown between constancy of visits during the first year (i.e. having an HIV primary care visit in each 3-month quarter) and survival. Another study examining care over a two-year period has found that mean increase from baseline CD4 counts was significantly greater among those with optimal retention (visits in all 4 six-month intervals) than among those with sub-optimal retention, and that mortality was higher among those with suboptimal retention.

In 2011, the HIV community saw the emergence of the HIV care continuum. This simple model outlines the sequential steps of medical care that people living with HIV go through from initial diagnosis to achieving the goal of viral suppression. The steps include diagnosis, linkage to care, retention in care, receipt of HIV antiretroviral therapy and viral suppression. This model has been incorporated into the National HIV/AIDS Strategy as it has focused all HIV prevention, care, and treatment efforts in the United States. As outlined in the model, all though there are five different steps, each step is dependent upon each other. For instance, you cannot become virally suppressed if you are not receiving HIV antiretroviral therapy or retained in medical care.

The most recent nationwide data from CDC dated 2014 estimates that although 86% of people living with HIV have been diagnosed, only 40% are engaged in care, 37% have been prescribed HIV antiretroviral therapy, and 30% have achieved viral suppression. Right now, we are at a very special time and place. Many states and large metropolitan areas across the United States have developed plans to end the HIV epidemic in the communities. These jurisdictions have used the HIV care continuum and its steps as the framework by which they have developed their plans.

Numerator Statement: Number of patients in the denominator who had at least one medical visit in each 6-month period of the 24-month measurement period with a minimum of 60 days between first medical visit in the prior 6-month period and the last medical visit in the subsequent 6-month period. (Measurement period is a consecutive 24-month period of time.)
Denominator Statement: Number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the first 6 months of the 24-month measurement period.
Denominator Exclusions: Patients who died at any time during the 24-month measurement period.

Measure Type: Process
Data Source: Paper Records
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Jan 07, 2013 **Most Recent Endorsement Date:** Jan 07, 2013

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☐ **Yes** ☒ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☒ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary or **Summary of prior review in 2012**

- The evidence focused on multiple studies examining the impact of treatment on preventing HIV transmission and monitoring of CD4 count and viral load.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

- The developer provided a diagram outlining the sequential steps of medical care that people living with HIV go through from initial diagnosis to ultimately achieving viral suppression.
- The rationale for this measure states that prompt linkage and retention in HIV care is related to improving patient outcomes. Retention in medical care among people living with HIV (PLWH) is associated with an increase in baseline CD4 count; those patients not retained in care experienced greater mortality than those who were retained in care.
- The evidence that supports this measure states that systematic monitoring of retention in care may include surveillance of visit adherence, gaps in care, and the number of visits during a specified period of time (note that this guideline is unrated).
- Another recommendation states that systematic monitoring of retention in care is recommended for PLWH (level AII).
- The developer also provides several guidelines on HIV care and treatment with varying levels of evidence.

Questions for the Committee:

For possible exception to the evidence criterion:

- *Does the committee agree that viral suppression is a related heath outcome performance measure?*
- *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable for medical visit frequency without empirical evidence?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure that HIV medical visit frequency is linked to improved outcomes?*

Guidance from the Evidence Algorithm

Process measure is evidence based (Box 3) → Evidence based on systematic review and grading of the body of empirical evidence (Box 7) → Possible related outcome measures (Box 10)→ No exception → Insufficient

Preliminary rating for evidence: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

RATIONALE: Although the developer provides multiple guidelines on HIV care, the guideline that supports the evidence is unrated and does not specify a specific time period to measure retention in care.

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer presented data from the HIV Research Network (a consortium of community and academic sites providing HIV care linked by a centralized Data Coordinating Center) on the number of patient's meeting the numerator criteria. The HIVRN is composed of 11 sites representing 4 major geographic divisions and of the insurance status and coverage types typical for the population in care. Data for 2011-2013 were not presented due to resource constraints.
- Patients were included in the numerator regardless of age, if they had a diagnosis of HIV and had a medical visit in the first 6 months of the measurement period. Patients who died were excluded.
- The mean performance rate was 66.7% in 2007-2008 and increased to 72.6% in 2014-2015.

	2014-2015 N=15,049	2009-2010 N=17,687	2008-2009 N=16,881	2007-2008 N=15,790
Minimum	55.1	50.1	42.5	47.1
Maximum	83.8	82.8	83.1	86.1
Mean	72.6	68.9	67.73	66.7
25th percentile	68.2	63.4	59.9	59.7
50th percentile	70.9	67.7	66.2	70.6
75th percentile	79.5	74.6	75.5	78.2

Disparities

- The developer presented client level performance scores for HIV medical visit frequency across four time periods. The table below shows disparities in HIV medical visit frequency among Hispanics, males and transgender and clients aged 18-29. Numbers are presented as percentages.

Demographic	2014-2015	2009-2010	2008-2009	2007-2008
African American/Caribbean	72.7	67.5	67.0	64.8
White, not Hispanic	75.2	67.9	65.8	67.3
Hispanic	67.9	73.9	72.9	71.2
Other	66.2	68.8	68.5	73.0
Male	69.9	68.	67.5	66.2
Female	76.0	69.8	68.4	68.2
Transgender	66.7	72.9	65.8	62.4
<18	88.7	87.8	87.3	87.2
18-29	62.9	56.8	54.2	53.3
30-49	67.5	66.4	66.0	64.6
50+	76.1	75.9	73.7	73.7

Questions for the Committee:

- *Does the Committee agree that there is a gap in performance on HIV medical visit frequency that warrants a national performance measure for continued endorsement?*
- *Is the Committee aware of additional disparities data related to HIV medical visit frequency?*

o *Does the data demonstrate an adequate problem for HIV medical visit frequency among people living with HIV?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

Evidence 1a.

*I agree that viral suppression is a related heath outcome performance measure.
I agree that it is acceptable (or beneficial) to hold providers accountable for medical visit frequency without empirical evidence.
It is not clear that there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure or that the time period of medical visit frequency is linked to improved outcomes.

*The developer provides rationale and recommendations from a panel of experts that the measure, as a proxy for assessing retention in care, is recommended for PLWH to monitor their progress along the HIV care continuum to achieve viral suppression. I am not aware of any new studies that change the evidence for this measure.

*Based on the algorithm the evidence submitted is insufficient given that there is no systematic review of the evidence specific to this measure.
Specific questions:
• Does the committee agree that viral suppression is a related heath outcome performance measure?
Yes, retention in care is a major predictor of viral suppression based on data reported by developer and numerous other studies.
• Does the SC agree that it is acceptable (or beneficial) to hold providers accountable for medical visit frequency without empirical evidence?
Yes, based on DHHS guidelines.
From guidelines: In patients on a stable, suppressive ARV regimen. Viral load should be repeated every 3 to 4 months (AIII) or as clinically indicated to confirm continuous viral suppression. Clinicians may extend the interval to 6 months for adherent patients whose viral load has been suppressed for more than 2 years and whose clinical and immunologic status is stable (AIII).
For the patient on a suppressive regimen whose CD4 count has consistently ranged between 300 and 500 cells/ mm3 for at least 2 years, the Panel recommends CD4 monitoring on an annual basis (BII). Continued CD4 monitoring for virologically suppressed patients whose CD4 counts have been consistently >500 cells/mm3 for at least 2 years may be considered optional (CIII). The CD4 count should be monitored more frequently, as clinically indicated, when there are changes in a patient's clinical status that may decrease CD4 count and thus prompt OI prophylaxis. Examples of such changes include the appearance of new HIV-associated clinical symptoms or initiation of treatment known to reduce CD4 cell count (e.g., interferon, chronic corticosteroids, or anti-neoplastic agents) (AIII). In patients who fail to maintain viral suppression while on ART, the Panel recommends CD4 count monitoring every 3 to 6 months (AIII) (see Virologic Failure and Suboptimal Immunologic Response section).
Table 13 – Strategies to improve adherence to ART and Retention in Care
Includes systematically monitor retention in care
• Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure that HIV medical visit frequency is linked to improved outcomes?
Yes, based on guideline review panel recommendations

Performance Gap 1b.

*There is a gap in performance on HIV medical visit frequency that warrants a national performance measure for continued endorsement. I'm not aware of other data.

*The developer provided performance data of the measure for 11 HIV care sites that represent geographic, insurance status and coverage types typical of people living with HIV who access care. The data show improvement in the measure over an 8 year time span and large variation across the sites. The data also show considerable variation among people of different race/ethnicities, gender, and age, suggesting disparities in routine (standard of care) receipt of medical care.

*Data for opportunity for improvement based on data provided is high.
Specific questions:
• Does the Committee agree that there is a gap in performance on HIV medical visit frequency that warrants a national performance measure for continued endorsement?
Yes, based on US data the majority of persons LWH are not retained in HIV care.
• Is the Committee aware of additional disparities data related to HIV medical visit frequency?
US data available that shows disparities in retention in care (link: https://www.cdc.gov/hiv/library/slidesets/)
• Does the data demonstrate an adequate problem for HIV medical visit frequency among people living with HIV?
Yes, based on data provided.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

- Paper records

Specifications:

- This measure is specified at the facility level in the clinician office/clinic.
- Patients are included in the numerator if they had at least one medical visit in each 6-month period of the 24-month measurement period with a minimum of 60 days between the first medical visit in the prior 6-month period and the last medical visit in the subsequent 6-month period.
- The denominator includes the number of HIV patients, regardless of age, with at least one medical visit in the first 6 months of the 24-month measurement period. Patients are excluded if they died at any time during the 24-month measurement period.
- The measure calculates a rate where a higher score is associated with better performance. The rate is calculated by dividing the numerator population by the denominator population and then multiplying by 100.

Questions for the Committee:

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if 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 enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- In the previous review of this measure, the developer conducted signal to noise testing to assess reliability.

Describe any updates to testing:

- Testing was not updated.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- The dataset included HIV Research Network data from the years 2007 – 2015 (data for 2011-2013 were not provided due to resource constraints). The HIVRN is a consortium of community and academic HIV providers care sites, linked by a centralized Data Coordinating Center. Testing data came from 13 sites and 17,687 patients participating in the HIVRN.
- The developer estimated reliability using a beta binomial model to assess the signal-to-noise ratio. The developer reports this model is appropriate for measuring the reliability since it calculates the ratio of signal to noise. Reliability scores fall from 0.0 to 1.0; where a reliability score of 1.0 implies that all variation is caused by real difference in performance across entities and 0.0 indicates that all variation is attributed to measurement error (i.e., noise).

Results of reliability testing

- Results showed a median reliability of 0.97, which the developer reported demonstrates good reliability.

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

Questions for the Committee:

- *No updated testing information is presented. The prior testing demonstrated good reliability. Does the Committee think there is a need to re-discuss and re-vote on reliability?*

If the Committee does not choose to re-vote, then a discussion may still be needed.

- *Is the measure score test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm

Precise specifications (Box 1) → Empirical reliability testing (Box 2) → Computed performance scores for measured entities (Box 4) → Signal-to-noise appropriate method used (Box 5) → High certainty that the performance scores are reliable based on the reliability statistic and scope of testing (# of measured entities and representativeness) (Box 6a) → High

Preliminary rating for reliability: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- At the previous review of this measure, the steering committee agreed that the measure met the scientific acceptability criteria. Face validity was used to establish measure validity but threats to validity were not assessed.

Describe any updates to validity testing:

- See updated face validity below.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☒ **Face validity only**

☐ **Empirical validity testing of the measure score**

Validity testing method:

- Face validity was established using a technical advisory panel. The panel was presented with current research in HIV care and treatment. Members then voted on the domains for the proposed measure based on importance, ability to assess quality of care, feasibility and use in quality improvement activities.
- NQF guidance states, "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.'

Validity testing results:

- The developer stated that "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." This is insufficient per NQF criteria.

Questions for the Committee:

- o *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- o *Do you agree with the score for this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients are excluded from the measure if they die during the measurement period, however the developer notes that patient mortality has declined over the years as a result of the development and dissemination of HIV antiretroviral therapy.
- The developer reports they were unable to assess the impact of exclusions due to constraints.

Questions for the Committee:

o *Are the exclusions consistent with the evidence?*

o *Are any patients or patient groups inappropriately excluded from the measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- As discussed above, the measure detects providers with better or worse than median performance scores. There is also a large difference between the minimum and maximum scores in each time period.
    - In 2014-2015, the mean performance for HIV medical visit frequency was 72.6%, up from 66.7% in 2007-2008. Providers in the 75th percentile had medical visit frequency rates at 79.5% in 2014-2015 compared to a rate of 68.2% for providers in the 25th percentile.

	2014-2015	2009-2010	2008-2009	2007-2008
# of Pts Included	15,049	17, 687	16, 881	15,790
Minimum	55.1%	50.1%	42.5%	47.1%
Maximum	83.8%	82.8%	83.1%	86.1%
Mean	72.6%	68.9%	67.73%	66.7%
25th percentile	68.2%	63.4%	59.9%	59.7%
50th percentile	70.9%	67.7%	66.2%	70.6%
75th percentile	79.5%	74.6%	75.5%	78.2%

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- Not applicable.

2b7. Missing Data

- The developer reports that missing data could not be assessed due to constraints.

Guidance from the Validity Algorithm

Specifications consistent with evidence (Box 1) → Relevant potential threats to validity assessed empirically assessed (Box 2) → Empirical validity testing was not conducted using the measure as specified (Box 3) → Face validity was not systematically assessed by recognized experts to determine agreement on whether the computed measure score from the measure as specified can be used to distinguish good and poor quality. Face validity focused on importance, ability to assess quality care, feasibility to implement measure, and use in quality improvement activities (e.g. ability to improve measure score). (Box 4) → Insufficient (highest eligible rating is MODERATE)

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

RATIONALE: Face validity was not systematically assessed by recognized experts to determine agreement on whether the computed measure score from the measure as specified can be used to distinguish good and poor quality per NQF criteria. Face validity focused on importance, ability to assess quality care, feasibility to implement measure, and use in quality improvement activities (e.g. ability to improve measure score).

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. Reliability Specifications

***All the data elements are clearly defined.
The calculation algorithm is clear.
It is likely this measure can be consistently implemented.

***Data elements and the measure itself are well defined.

*Are all the data elements clearly defined? Are all appropriate codes included?
Yes
• Is the logic or calculation algorithm clear?
Yes

• Is it likely this measure can be consistently implemented?
Yes. We have been using this measure locally and it can be consistently implemented across different provider types.

2a2. Reliability Testing
*****There is no need to re-discuss and re-vote on reliability.
The measure score test sample is adequate to generalize for widespread implementation.
The results demonstrate sufficient reliability so that differences in performance can be identified.

*****Results provided by the developer show a high reliability overall and for each of the 11 sites used to assess the measure, demonstrating sufficient reliability to detect differences in performance.

***** No updated testing information is presented. The prior testing demonstrated good reliability. Does the Committee think there is a need to re-discuss and re-vote on reliability?
The testing information shows high reliability.

2b1. Validity Specifications
*The specifications are consistent with the evidence.

*Specifications are consistent with the evidence

*Face validity only

2b2. Validity Testing
*****The results demonstrate sufficient validity so that conclusions about quality can be made.
Though guidelines agree, I do not see evidence that the score for this measure as specified is an indicator of quality.

*****Face validity was established through a technical advisory panel who voted on the measure based on importance, ability to assess quality of care and feasibility of use to inform improvements, but they did not determine agreement on whether or not the computed measure score could be used to distinguish good or poor quality as required by NQF criteria.

*Do the results demonstrate sufficient validity so that conclusions about quality can be made?
Face validity only. Unclear why the developer has not tested validity since approval in 2013.
• Do you agree with the score for this measure as specified is an indicator of quality
Yes

2b3-7. Threats to Validity
*2b.3 It's not possible to comment on the merit of exclusions, but they are logical.

*Exclusions are consistent with the evidence and it is unlikely that patients are inappropriately excluded from the measure. The measure identifies meaningful differences between the 11 sites for which data were obtained as well as differences over an 8 year time period.

*Face validity only.
Analyses of data provided indicate that the measure identifies meaningful differences by year and by provider.

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The developer reports that data elements are generated or collected by and used by healthcare personnel during the provision of care.
- The developer reports that all data elements are in defined fields in electronic health records, and that data are readily available within patient health records. Data are provided annually to the HIVRN.
- There are no fees, licensing, or other requirements to use the measure.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

*The developer states that all data elements are collected, generated, and used as part of routine delivery of care

*Feasability is rated high.
Specific questions:
• Are the required data elements routinely generated and used during care delivery?
Yes.
• Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
Yes

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No** ☐ **UNCLEAR**

Accountability program details

- Physician Quality Reporting System and Value Based Modifier
  Sponsor: Federal government
  Geographic area: Nationwide
  Accountable entities: Physicians and practitioners
  Patients: Unknown

- Merit-Based Incentive Payment System
  Sponsor: Federal government
  Geographic area: Nationwide
  Accountable entities: Physicians, Physician Assistant, Nurse Practitioner, and Clinical Nurse Specialist
  Patients: Unknown

- This measure is used in the Ryan White HIV/AIDS program which provides grants to over 600 recipients and their providers. The RWHAP serves approximately 316,000 patients.

Improvement results

- The developer reports that medical visit frequency performance has improved over time.
  - Based on HIVRN data of over 15,000 patients, performance has increased from 66.7% in 2007-2008 to 72.6% in 2014-2015.

Unexpected findings (positive or negative) during implementation

- The developer reports that since the development of this measure, it has been adopted by the Centers for Medicare & Medicaid Services measurement programs and selected as a core HIV indicator by the Secretary of the Department of Health and Human Services.
- National learning collaborative's for HIV/AIDS quality improvement activities have also used the measure for RWHAP grant recipients and sub-recipients.

Potential harms

- The developer did not identify any potential harms in the testing of this measure.

Vetting of the measure

- During the initial development of the measure, the developer reports that formal feedback was gathered.
- The developer reports that the measure is reviewed annually for clinical relevance, change in scientific acceptability, and consistency with guidelines.

Feedback:

- The developer reports that RWHAP grant recipients have provided positive and supportive feedback for this measure. RWHAP grant recipients have encouraged further stratification, dissemination methods, and graphical presentations.
- Additional feedback notes the encouragement of alignment of measure details (e.g. numerator, denominator, exclusions) across related performance measures and measure programs in order to reduce burden.

***Questions for the Committee**:*

- o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*
- o *How has the measure been vetted in real-world settings by those being measure or others?*

Preliminary rating for usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4. Usability and Use

*"The measure is being used for the Physician Quality Reporting Systems and Value Based Modifier, the Merit-Based Payment System, and the HRSA-Ryan White HIV/AIDS Program (RWHAP).

The developers received feedback during the initial development of the measure, and reports that RWHAP grant recipients have provided positive and supportive feedback. RWHAP recipients also suggested that the measure elements be aligned across related performance measures.

While the measure has been vetted in the real-world, it is not apparent from the literature and federal data reports, how frequently it is used by providers, and other programs. This measure does not align with the same time period as the ""Gap in Visits"", Viral Suppression"" and ""ART prescription"" NQF HIV-related measures, thus making it difficult to align measures along the continuum of care for a specified time period. Furthermore, results from this measure are not described in the HRSA Annual RSR report that describes indicators of HIV Care, but they instead use the definition of retention used by the CDC to measure progress along the HIV continuum of care to viral suppression. "

* How can the performance results be used to further the goal of high-quality, efficient healthcare?
Good question
• Do the benefits of the measure outweigh any potential unintended consequences?
Yes. No potential unintended consequences.
• How has the measure been vetted in real-world settings by those being measure or others?
Local data also supports.

Criterion 5: Related and Competing Measures

Related or competing measures

- The following measures are listed as related or competing:
  - 2080 Gap in HIV Medical Visits – same population but different measurement periods and focuses on patients that did not get a visit.
  - 2082 HIV viral suppression
  - 2083 Prescription of HIV Antiretroviral Therapy
  - 3211 Prescription of HIV Antiretroviral Therapy (newly submitted eMeasure)
  - 3210 HIV viral suppression (newly submitted eMeasure)
  - 3010 HIV Medical Visit Frequency
  - 0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis – related population only
  - 0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis – related population only

Harmonization

- The developer notes that this measure is harmonized with the first 6 measures listed above. For these 6 measures, the target population is the same (i.e., people living with HIV) however the measure focus is different.
- The developer plans to harmonize with #0405 and #0409. At this time, #0405 and 0409 have been granted a deferral for maintenance of endorsement. There are no additional steps the developer needs to take to harmonize this measure with #0405 or #0409 since the measure focus is different (HIV patients receiving PCP prophylaxis and those screened for STDs).

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a <u>candidate</u> for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

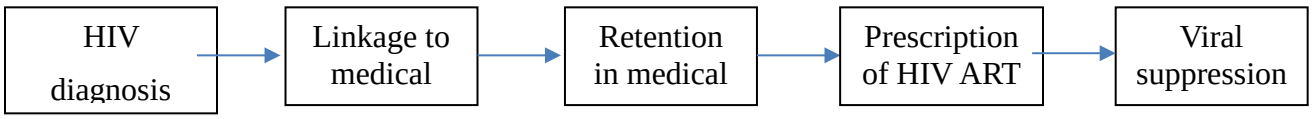
RATIONALE IF NOT ELIGIBLE:

This measure is not eligible for Endorsement + designation since the developer did not perform empirical validity testing of the measure score.

Pre-meeting public and member comments

- Representing the HIV Quality of Care Advisory Committee and Consumer Advisory Committee of the NY State Department of Health AIDS institute we would advise that this measure be dropped based on the variation of expected frequency of visits for patients based on their viral load suppression status. Frequency measures suggest that a rigid spacing of intervals of visits can be universally applied which is no longer the global standard of care - even in resource limlited settings where differentiated care models are promoted by WHO and the Global Fund. The measure is not as useful at clinic level for improvement as missed visit measures, based on extensive research led by Mugavero among others.

1a.12 LOGIC MODEL



Although the above diagram outlines the sequential septs of medical care that people living with HIV go through form initial diagnosis to achieving the goal of viral suppression (also known as the HIV care continuum). For some patients, this is a linear path with sustained viral suppression for many years. For other patients, there may be years between diagnosis and linkage. Yet still for others, retention in medical care is not consistent, which results in missed visits, no prescription for or adherence to HIV antiretroviral therapy (ART), and lack of viral suppression.

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

Regularly attending medical visits (retention) is paramount to monitoring patient's health status, screenings, and laboratory values. Providers need this information to make an informed decision in order to prescribe HIV antiretroviral therapy (ART). ART reduces HIV-associated morbidity and mortality by maximally inhibiting HIV replication (as defined by achieving and maintaining plasma HIV RNA (viral load) below levels detectable by commercially available assays). Durable viral suppression improves immune function and quality of life, lowers the risk of both AIDS-defining and non-AIDS-defining complications, and prolongs life. Emerging evidence also suggests that additional benefits of ART-induced viral load suppression include a reduction in HIV-associated inflammation and possibly its associated complications.

In 2011, the HIV community saw the emergence of the HIV care continuum. This simple model outlines the sequential steps of medical care that people living with HIV go through from initial diagnosis to achieving the goal of viral suppression. The steps include diagnosis, linkage to care, retention in care, receipt of HIV antiretroviral therapy and viral suppression. This model has been incorporated into the National HIV/AIDS Strategy as it has focused all HIV prevention, care, and treatment efforts in the United States. As outlined in the model, all though there are five different steps, each step is dependent upon each other. For instance, you cannot become virally suppressed if you are not receiving HIV antiretroviral therapy or retained in medical care.

The most recent nationwide data from CDC dated 2014 estimates that although 86% of people living with HIV have been diagnosed, only 40% are engaged in care, 37% have been prescribed HIV antiretroviral therapy, and 30% have achieved viral suppression.

Right now, we are at a very special time and place. Many states and large metropolitan areas across the United States have developed plans to end the HIV epidemic in the communities. These jurisdictions have used the HIV care continuum and its steps as the framework by which they have developed their plans.

In closing, the measures we have put forth are in alignment with the HIV care continuum. We see these measures as a suite – each important as individual measures, but work together as a suite to improve health outcomes for people living with HIV in the United States.

1b.1. Briefly explain the rationale for this measure (e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)

Prompt linkage to, and sustained retention in, HIV medical care have been clearly shown to maximize patient outcomes. Retention in medical care among people living with HIV is associated with a significantly greater mean increase in baseline CD4 count. Consequently, mortality was higher among those with suboptimal retention.

Poor retention in care during the first year of outpatient medical care is associated with delayed or failed receipt of antiretroviral therapy, delayed time to virologic suppression and greater cumulative HIV burden, increased sexual risk transmission behaviors, increased risk of long-term adverse clinical events, and low adherence to antiretroviral therapy.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services, Accessed November 18, 2016: http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf

World Health Organization (WHO). (2016). Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach. Accessed November 18, 2016: http://apps.who.int/iris/bitstream/10665/208825/1/9789241549684_eng.pdf?ua=1

International Advisory Panel on HIV Care Continuum Optimization (IAPAC). (2015). IAPAC Guidelines for Optimizing the HIV Care Continuum for Adults and Adolescents. Accessed November 18, 2016. http://www.iapac.org/uploads/JIAPAC-IAPAC-Guidelines-for-Optimizing-the-HIV-Care-Continuum-Supplement-Nov-Dec-2015.pdf

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Panel on Antiretroviral Guidelines for Adults and Adolescents: (unrated)

- The critical elements of adherence go hand in hand with linkage-to-care and retention in care. A recently released guideline provides a number of strategies to improve entry and retention in care and adherence to therapy for HIV infected patients. As with adherence monitoring, research advances offer many options for systematic monitoring of retention in care that may be used in accordance with local resources and standards. The options include surveillance of visit adherence, gaps in care, and the number of visits during a specified period of time. (page K-4)
- In addition to maintaining high levels of medication adherence, attention to effective linkage to care, engagement in care, and retention in care is critical for successful treatment outcomes. To foster treatment success, there are interventions to support each step in the cascade of care, as well as guidance on systematic monitoring of each step in the cascade. (page K-4)
- Where youth services are available, they may be helpful to consider as one approach to enhancing HIV care engagement and retention among adolescents. Regardless of the setting, expertise in caring for adolescents is critical to creating a supportive environment for engaging youth in care. (I-9)

World Health Organization:

Section 6. 5 Retention in care (page 251)

- Programmes should provide community support for people living with HIV to improve retention in HIV care (strong recommendation, low-quality evidence).
- The following community-level interventions have demonstrated benefit in improving retention in care:
    - package of community based interventions (children low-quality and adults very low-quality evidence)
    - adherence clubs (moderate-quality evidence)
    - extra care for high-risk people (very low-quality evidence).

Section 6.7 Frequency of clinical visits and medical pick-up (page 259)

- Less frequent clinical visits (3–6 months) are recommended for people stable on ART (strong recommendation, moderate-quality evidence)

- Less frequent medication pickups (3-6 months) are recommended for people stable on ART (strong recommendation, low-quality evidence)

IAPAC on HIV Care Continuum Optimization: (page 6)

23. Systematic monitoring of retention in HIV care is recommended for all patients. (A II)
23a. Retention in HIV care should be considered as a quality indicator. (B III)
23b. Measuring retention in HIV care using electronic health record and other health system data is recommended. (BII)
23c. Use of clinic databases/surveillance systems for HIV clinical monitoring and population-level tracking is recommended. (B II)
26. Patient education about and offering support for medication adherence and keeping clinic appointments are recommended. (A I)
28. Proactive engagement and reengagement of patients who miss clinic appointments and/or are lost to follow-up, including intensive outreach for those not engaged in care within 1 month of a new HIV diagnosis, is recommended. (B II)
28a. Case management to retain PLHIV in care and to locate and reengage patients lost to follow-up is recommended. (B II)
28b. Transportation support for PLHIV to attend their clinic visits is recommended. (B II)

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:
Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents and Recommendations for Use of Antiretroviral Drugs in Pregnant HIV-1-Infected Women for Maternal Health and Interventions to Reduce Perinatal HIV Transmission in the United States

Basis for Recommendations
Recommendations in these guidelines are based upon scientific evidence and expert opinion. Each recommended statement includes a letter (A, B, or C) that represents the strength of the recommendation and a Roman numeral (I, II, or III) that represents the quality of the evidence that supports the recommendation (see Table 2).

Table 2. Rating Scheme for Recommendations

Strength of Recommendation	Quality of Evidence for Recommendation
A: Strong recommendation for the statement **B:** Moderate recommendation for the statement **C:** Optional recommendation for the statement	**I:** One or more randomized trials with clinical outcomes and/or validated laboratory endpoints **II:** One or more well-designed, non-randomized trials or observational cohort studies with long-term clinical outcomes **III:** Expert opinion

International Advisory Panel on HIV Care Continuum Optimization; IAPAC Guidelines for Optimizing the HIV Care Continuum for Adults and Adolescents.

Strong (A) = Almost all patients should receive the recommended course of action.
Moderate (B) = Most patients should receive the recommended course of action. However, other choices may be appropriate for some patients.
Optional (C) There may be consideration for this recommendation based on individual patient circumstances. Not recommended routinely.

Quality of the Body of Evidence and its Interpretation:

Excellent (I) = Randomized control trial (RCT) evidence without important limitations; overwhelming evidence from observational studies
High (II) = RCT evidence with important limitations; strong evidence from observational studies
Medium (III) = RCT evidence with critical limitations; observational study without important limitations

Low (IV) = Other evidence, including extrapolations from bench research, usual practice, expert opinion, consensus guidelines; observational study evidence with important or critical limitations

World Health Organization. Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection Recommendations for a public health approach - Second edition.

The strength of a recommendation can be either strong or conditional.
Process of guideline development This edition of the guidelines was revised in accordance with procedures established by the WHO Guidelines Review Committee. New clinical and operational recommendations in the guidelines are based on the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) approach to reviewing evidence. Modelling, expert consultations and country case studies have all strongly informed the guidelines. The process has also identified key gaps in knowledge that will help to guide the future HIV research agenda. A strong recommendation is one for which there is confidence that the desirable effects of adherence to the recommendation clearly outweigh the undesirable effects.
A conditional recommendation is one for which the Guideline Development Group concludes that the desirable effects of adherence to the recommendation probably outweigh the undesirable effects or are closely balanced, but the Groups are not confident about these trade-offs in all situations. At implementation, monitoring and rigorous evaluation is needed to address these uncertainties, which are likely to provide new evidence that may change the calculation of the balance of trade-offs and to suggest how to overcome any implementation challenges.

Quality of evidence Definition
Table 1.1. GRADE quality of evidence

Quality of evidence	Definition
High	We are very confident that the true effect lies close to that of the estimate of the effect
Middle	We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of effect, but there is a possibility that it is substantially different
Low	Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect
Very low	We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of the effect

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)
All grade and definitions noted in 1a.4.3

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**
Citations noted in 1a.4.1

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?
X☐ Yes → ***complete section 1a.7***
☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

MVF_evidence_NQF.docx,MVF_submission-636179047812919962.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

Yes

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*
IF a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)
IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.

Poor retention in care during the first year of outpatient medical care is associated with delayed or failed receipt of antiretroviral therapy, delayed time to virologic suppression and greater cumulative HIV burden, increased sexual risk transmission behaviors, increased risk of long-term adverse clinical events, and low adherence to antiretroviral therapy. Early retention in HIV care has been found to be associated with time to viral load suppression and 2-year cumulative viral load burden among patients newly initiating HIV medical care (8). In this study, each "no show" clinic visit conveyed a 17% increased risk of delayed viral load suppression. A dose- response relationship has been shown between constancy of visits during the first year (i.e. having an HIV primary care visit in each 3-month quarter) and survival. Another study examining care over a two-year period has found that mean increase from baseline CD4 counts was significantly greater among those with optimal retention (visits in all 4 six-month intervals) than among those with sub-optimal retention, and that mortality was higher among those with suboptimal retention.

In 2011, the HIV community saw the emergence of the HIV care continuum. This simple model outlines the sequential steps of medical care that people living with HIV go through from initial diagnosis to achieving the goal of viral suppression. The steps include diagnosis, linkage to care, retention in care, receipt of HIV antiretroviral therapy and viral suppression. This model has been incorporated into the National HIV/AIDS Strategy as it has focused all HIV prevention, care, and treatment efforts in the United States. As outlined in the model, all though there are five different steps, each step is dependent upon each other. For instance, you cannot become virally suppressed if you are not receiving HIV antiretroviral therapy or retained in medical care.

The most recent nationwide data from CDC dated 2014 estimates that although 86% of people living with HIV have been diagnosed, only 40% are engaged in care, 37% have been prescribed HIV antiretroviral therapy, and 30% have achieved viral suppression.

Right now, we are at a very special time and place. Many states and large metropolitan areas across the United States have developed plans to end the HIV epidemic in the communities. These jurisdictions have used the HIV care continuum and its steps as the framework by which they have developed their plans.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis*. (This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

Please see attachment "MVF submission form" for formatted data.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
N/A

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*
Please see attachment "MVF submission form" for formatted data.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4**
N/A

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Infectious Diseases (ID) : HIV/AIDS

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

De.7. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Medical_visit_frequency_data_dictionary.pdf

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.
No

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.
None

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Number of patients in the denominator who had at least one medical visit in each 6-month period of the 24-month measurement period with a minimum of 60 days between first medical visit in the prior 6-month period and the last medical visit in the subsequent 6-month period. (Measurement period is a consecutive 24-month period of time.)

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
To be included in the numerator, patients must have had at least one medical visit in each 6-month period of the 24-month measurement period with a minimum of 60 days between first medical visit in the prior 6-month period and the last medical visit in the subsequent 6-month period.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the first 6 months of the 24-month measurement period.

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
To be included in the denominator, patients must meet all of the following conditions/events:
1. Patients of any age during the measurement period
2. Patients without a date of death during the 24-month measurement period
3. Patients diagnosed with HIV during the first 3 months of the 24-month measurement period or prior to the measurement period
4. Patients who had at least one medical visit in the first 6 months of the 24-month measurement period

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients who died at any time during the 24-month measurement period.

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
Patients with a date of death during the measurement period.

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*
Not applicable

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
No risk adjustment or risk stratification
If other:

S.12. **Type of score:**
Rate/proportion
If other:

S.13. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
1. Identify the individuals who satisfy all specific criteria for inclusion in the denominator: 1.) diagnosed with HIV during the first 3 months of the 24-month measurement period or prior to the 24-month measurement period; 2.) did not have a date of death during the 24-month measurement period; and 3.) had at least one medical visit in the first 6 months of the 24-month measurement period. The individuals who met these three criteria are the denominator population.
2. Identify the individuals from the denominator population who meet the criterion for inclusion in the numerator: must have had at least one medical visit in each 6-month period of the 24-month measurement period with a minimum of 60 days between first medical visit in the prior 6-month period and the last medical visit in the subsequent 6-month period.
3. Calculate the rate by dividing the numerator population by the denominator population and multiply by 100.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable; not based on a sample.

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
This measure is not based on a survey or instrument.

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Paper Records

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Electronic or paper records

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.20. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Clinician Office/Clinic
If other:

S.22. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
This is not a composite measure.

2. Validity – See attached Measure Testing Submission Form
MVF_testing.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)
Yes

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)
No

2.3 For maintenance of endorsement
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

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).

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)

What were the statistical results of the analyses used to select risk factors?

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)
No - This measure is not risk-adjusted

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2079
Measure Title: HIV medical visit frequency
Date of Submission:
Type of Measure:

☐ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐Intermediate Clinical Outcome	☐ Cost/resource
☒ 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:**
☒ 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:	☐ other:

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, 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
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

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, 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*)
☒ **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**

☒ **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

2. 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).

_____

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 ot 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.

2b4.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
☐ **Statistical risk model with risk factors**
☐ **Stratification by risk categories**
☐ **Other,**

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

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

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.
ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance 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.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).
Not applicable.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/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, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.
Data availability: The data used for testing and operational use of this measure are readily available within patient health records and provided annually to HIVRN.

Missing date: We were not able to assess for missing data in this submission due to constraints when working with the HIVRN.

Time and frequency of data collection: As noted previously, all variables to calculate this measure are contained in a patient health record in a structured field. These data are routinely collected in the provision of care to people living with HIV. Because the availability of data, sampling is not performed.

Patient confidentiality: The data used in the testing of this measure are deidentified/striped of personally identifiable information prior to submitting.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
No fees, licensing, or other requirements to use any aspect of the measure.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting	Public Health/Disease Surveillance Ryan White HIV/AIDS Program https://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio Payment Program PQRS

	https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html?redirect=/pqri Quality Improvement (external benchmarking to organizations) Ryan White HIV/AIDS Program https://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio Quality Improvement (Internal to the specific organization) Ryan White HIV/AIDS Program https://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio

4a.1. **For each CURRENT use, checked above (update for <u>maintenance of endorsement</u>), provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

Ryan White HIV/AIDS Program
Sponsor: Federal government
Geographic area: Nationwide
Accountable entities: Approximately 600 Ryan White HIV/AIDS Program grant recipients and their providers
Patients: Approximately 316,000 patients

Physician Quality Report System and Value Based Modifier
Sponsor: Federal government
Geographic area: Nationwide
Accountable entities: Physicians and practitioners
Patients: Unknown

Merit-Based Incentive Payment System
Sponsor: Federal government
Geographic area: Nationwide
Accountable entities: Physicians, Physician Assistant, Nurse Practitioner, and Clinical Nurse Specialist
Patients: Unknown

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- 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 timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
N/A

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Medical visit frequency is a measurement of retention in HIV medical care and specifically geared towards longer term retention. Performance has been improving over time. Based on the HIVRN data, representing over 15,000 patients annually, performance has increased from 66.7% in 2007-2008 to 72.6% in 2014-2015. Many, but not all of the demographic groups and subpopulations have seen improvements in the medical visit frequency measure.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**

The adoption and use of this measure has continued to spread since the initial development of this measure. This measure has been adopted by Centers for Medicare and Medicaid measurement programs, Department of Health and Human Service Secretary as a one of the core HIV indicators, countless outpatient/ambulatory care settings, and health departments. National learning collaborates have used this measure to focus the improvement efforts of grant recipients and subrecipients. Additionally, retention is the final and goal of the five stages of the HIV care continuum.

4c.2. **Please explain any unexpected benefits from implementation of this measure.**

N/A

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

This measure has been used in national quality improvement campaigns, learning collaborative, and learning exchange. Participants commit to using this measure, reporting performance scores and disparity stratifications, and developing quality improvement projects based on this measure. Performance scores and disparity stratification data are shared with participants in order to benchmark performance.

HRSA is releasing a quality module where grant recipients can voluntarily report numerator, denominator, and performance scores for a portfolio of measures. Grant recipients will be able to benchmark their performance based on a number of patient demographic and organizational factors. This measure will be included in the measure portfolio.

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

For the national quality improvement campaign, data were collected and aggregated from participants across the United States every other month. Reports were developed and released based on a number of organizational factors (type of funding, location, etc.). Reports included data tables and spark lines and available on a public website and presented in public, national webinars. Similar efforts were employed for the learning collaborative and learning exchange.

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.

Antidotal feedback has been received regarding the use of performance measures, collection of data, and dissemination of reports from participating Ryan White HIV/AIDS Program grant recipients. All of the feedback was positive, supportive, and encouraged further stratification, dissemination methods, and graphical presentations. Feedback was incorporated in dissemination efforts based on feasibility and resource availability.

4d2.2. **Summarize the feedback obtained from those being measured.**

See 4d2.2

4d2.3. **Summarize the feedback obtained from other users**

Antidotal feedback encouraged continual alignment of measure details (e.g. numerator, denominator, exclusions, etc.) across performance measures and measure programs in order to reduce burden.

4d.3. Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

During the initial development of the measure, formal feedback was gathered. The measures were modified during the development phase and have not been modified since. A concerted effort was made to develop a measure that would likely stand the test of time from a scientific, clinical, and patient perspective. On an annual basis, the measure is review for clinical relevance, change in scientific acceptability, and consistency with guidelines. This measure has not been modified as a result of the annual reviews. Additionally, this measure is used by a number of measurement programs and strategies. Each of those programs require a separate annual review. No modifications have been made for those programs.

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

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

0403 : HIV/AIDS: Medical Visit

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis
0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis
2080 Gap in HIV Medical Visits
2082 HIV viral suppression
2083 Prescription of HIV Antiretroviral Therapy
3211 Prescription of HIV Antiretroviral Therapy
3210 HIV viral suppression
3010 HIV Medical Visit Frequency

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?

Yes

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

Harmonized with all measures except 405 and 409. Plans to harmonize with 405 and 409.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed 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.)
This measure does not have a competing measure.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Health Resources and Services Administration - HIV/AIDS Bureau
Co.2 **Point of Contact:** Marlene, Matosky, mmatosky@hrsa.gov, 301-443-0798-
Co.3 **Measure Developer if different from Measure Steward:** Health Resources and Services Administration - HIV/AIDS Bureau
Co.4 **Point of Contact:** Marlene, Matosky, mmatosky@hrsa.gov, 301-443-0798-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The work group members determined the measure concepts, identified the data elements, voted on the final measures, and assessed the face validity of the measures.
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, HRSA 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
Karam Mounzer, Philadelphia Fight, Philadelphia, PA
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

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2011

Ad.3 **Month and Year of most recent revision:** 05, 2016

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

Ad.5 **When is the next scheduled review/update for this measure?** 05, 2017

Ad.6 **Copyright statement:** None

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None