NATIONAL QUALITY FORUM

MEASURE WORKSHEETS

- 2079 HIV Medical Visit Frequency
- 3209 HIV Medical Visit Frequency eMeasure
- 2080 Gaps in Medical Visits
- 2082 HIV Viral Load Suppression
- 3210 HIV Viral Load Suppression eMeasure
- 2083 Prescription of HIV Antiretroviral Therapy
- 3211 Prescription of HIV Antiretroviral Therapy eMeasure
- 3215 Adult Inpatient Risk Adjusted Sepsis Mortality
- 0500 Severe Sepsis and Septic Shock: Management Bundle

NATIONAL QUALITY FORUM

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?*
- *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:

- *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:

Do the results demonstrate sufficient validity so that conclusions about quality can be made?

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:

- *Are the exclusions consistent with the evidence?*
- *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**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*
- *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 candidate 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** X **No**

RATIONALE IF NOT ELIGIBLE:

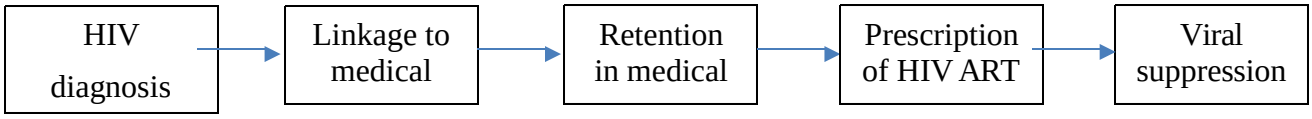
This measure is not eligible for Endorsement + designation since the developer did not perform empirical validity testing of the measure score.

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.

Measure Title: HIV Medical Visit Frequency

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

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

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: Click here to enter a date
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: Click here to describe	☐ other: Click here to describe

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

We utilized the multisite HIV Research Network (HIVRN), a consortium of community and academic HIV providers care sites, linked by a centralized Data Coordinating Center (DCC). The HIVRN has 11 participating treatment sites (10 adolescent/adult sites and 1 pediatric site). The sites are representative of both academic and community-based HIV care; of the 4 major geographic divisions of the U.S. of the demographic diversity of HIV infection across the U.S. and of the insurance status and coverage types typical of the population in care. The measurement periods included calendar years 2007-2008, 2008-2009, 2009-2010, 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 <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.26*)**	**Measure Tested at Level of:**

☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

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___.

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.1.1 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 <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> 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 <u>and</u> 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 <u>or</u> 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 _____

2b4.6. 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

NATIONAL QUALITY FORUM

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 #: 3209
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 within 24 months 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: Patients who had at least one medical visit in each 6-month of a consecutive consecutive 24 month 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.

Denominator Statement: : Patients, regardless of age, diagnosed with HIV during the first 3 months of the year preceding the measurement period or prior to the measurement period with at least one medical visit in the first 6 months of the year preceding the measurement period. The target population for this measure is all people living with HIV. **Denominator Exclusions:** Patients who died at any time during the measurement period or the 12 months preceding the measurement period.
Measure Type: Process **Data Source:** Electronic Health Record (Only) **Level of Analysis:** Facility

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

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.

This measure is the new eMeasure version of NQF #2079. The information provided for Evidence and Opportunity for Improvement is identical to that submitted for NQF #2079. Measure #2079 will be discussed first – the ratings for evidence and opportunity for improvement will automatically be assigned to this eMeasure without further discussion.

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

- 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 experience 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).
    - Measuring retention in HIV care using electronic health record and other health system data is recommended (BII)
- The developer also provides several other 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.

- There is no performance data available from this eCQM. However, 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.
- Summary statistics for the proportion of of 2014-2015 patients meeting the numerator are provided below. The performance rate was 66.7% in 2007-2008 and increased to 72.6% in 2014-2015. The table is found here.

	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 from the paper based measure, #2079. The table below shows disparities in HIV medical visit frequency among Hispanics, males and transgender and clients aged 18-29.

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:

- *Without data from the eMeasure as specified, do you agree that there is a quality problem with retaining patients in care?*
- *Is the Committee aware of additional disparities data related to HIV medical visit frequency?*
- *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)

1a. Evidence

*I don't understand the denominator statement.
I agree that viral suppression is a related heath outcome performance measure
It is acceptable (or beneficial) to hold providers accountable for medical visit frequency without empirical evidence
There are guidelines, but no clear 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. Also, the time frames are unsupported."

*See comments from NQF#2079
*
Identical to #2079

1b. Performance Gap

*I agree that there is a quality problem with retaining patients in care
The data demonstrate a problem for HIV medical visit frequency among people living with HIV

*No performance data available from this eCQM, but see comments on NQF#2079 for comments on performance gap using HIVRN data.

*Identical to #2079

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): Electronic health record only. This is an eMeasure.

Specifications:

- HQMF specifications for this eMeasure are included in the document set on SharePoint. See eMeasure Technical Review below.
- The level of analysis is at the facility level.
- The numerator includes patients who had at least one medical visit in each 6-month of a consecutive 24 month 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 patients, regardless of age, diagnosed with HIV during the first 3 months of the year preceding the measurement period or prior to the measurement period with at least one medical visit in the first 6 months of the year preceding the measurement period.
- Patients are excluded if they died at any time during the measurement period or the 12 months preceding the measurement period.
- The value sets needed to calculate the numerator and denominator are included in the specifications.
- The calculation algorithm is included.

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?*

eMeasure Technical Advisor(s) review (if not an eMeasure, delete this section):

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously;
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with measure developer indicates that the measure logic is feasible based on assessment by EHR vendors

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.

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 used for testing included 64 synthetic patients created in the Bonnie testing system simulating the year 2012. The developer tested the following data elements using the Bonnie testing tool to evaluate the measure logic:
    - Patient name
    - Date of birth
    - Race
    - Ethnicity
    - Gender
    - Payer
    - Diagnosis
    - Encounters
- The patient's bundle demographics were designed to mimic the HIV/AIDS population, specifically drawing from the patient characteristicds collected via the Ryan White HIV/AIDS Program Services Report (RSR).
- Data element validity testing was performed and will count for data element reliability – see validity testing section below.
- The developer provided reliability results from the paper based version of this measure (#2079) and stated, "Currently, there is no performance data available to test the eCQM. However, the chart-abstracted version of this measure has been in use in national quality reporting programs since as early as 2010."

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results from the Bonnie tool demonstrate sufficient reliability so that differences in performance can be identified?*
- *Do you agree that the reliability test results of the eMeasure will be comparable to the paper based measure (#2079)?*

Guidance from the Reliability Algorithm Precise specifications (Box 1) →Empirical reliability testing (Box 2) →Empirical validity testing of patient-level data (Box 3) →Refer to validity testing of patient-level data elements using Bonnie tool (Box 10 of the Validity algorithm) →Method appropriate for legacy eMeasures (Box 11) → Moderate (Moderate is the highest possible rating)

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:

- *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.

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

Validity testing method:

- The Bonnie testing tool, with 64 synthetic patient records were used to test the measure logic and data elements.
  - For each synthetic patients, an expected result was assigned to reflect an expected result of the measure. The synthetic patients were then run against the HQMF output loaded into Bonnie, which "calculates" a measure result for each patients and evaluates it against the expected result.
  - A patient is considered to pass Bonnie testing when the expected result matches the "calculated" result.
- The following testing was completed on the synthetic patients
  - 100% logic coverage: The bundle of synthetic patients collectively includes all data elements and conditions that are specified within the measure logic.
  - Edge case testing: Data elements that test the upper or lower boundary of measure logic conditions.
  - Negative testing: Use of test cases that do not evaluate positively against the measure logic but are otherwise clinically relevant and realistic.
- The developer used references cited within the chart abstracted measure specifications to ensure the eCQM logic maintained alignment with the clinical intent of the chart abstracted measure.
- In addition to Bonnie testing, the measure specifications were reviewed independently by three eCQM experts to confirm the logic was syntactically correct, using appropriate and current versions of the eCQM standards and terminologies, and consistent with the intent of the chart-abstracted measure.

Validity testing results:

- The testing results from the Bonnie tool reached 100% coverage and confirmed there was a test case for each pathway of logic (negative and positive test cases).
- The measure had a 100% passing rate which confirmed that all the test cases performed as expected.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results from the Bonnie tool demonstrate sufficient validity so that conclusions can be made about quality?*
- *Do you agree that the results of the eMeasure will be comparable to the chart-abstracted measure (#2079)?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- This measure has one exclusion – patient death during the measurement period. The developer reports that the exclusion was tested similarly to other criteria using synthetic patients in Bonnie. When the exclusion element was present, the patients were correctly excluded from the measure. In the absence of the exclusion element, cases were not excluded from the measure.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *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 in the paper based version (#2079), the measure detects providers with better or worse than median performance scores. There is 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 the Committee agree the e-Measure will demonstrate similar results to the chart-abstracted measure?*

2b6. Comparability of data sources/methods:

- Not applicable

2b7. Missing Data

- Per the developer, "The HQMF standard specifies that if data are unknown or missing, they shall fail the criterion. This constraint embodies the notion that absence of evidence is evidence of absence, i.e. data not present in a structured field from which the measure draws will not be considered for measure calculation. In certain cases, missing data may have no impact on the measure outcome for a given patient. For example, a data element used in a series of OR statements will not impact the measure outcome if another data element in the OR statement is present and meets all other defined constraints."
- All Bonnie synthetic patients with missing data performed according to the HQMF standard specification and as expected.

Guidance from the Validity Algorithm Specifications consistent with evidence (Box 1)→Some threats to validity addressed (Box 2) →Empirical validity testing (Box 3) → Face validity testing (Box 4) and empirical testing of data elements using Bonnie tool (Box 10) →Method appropriate for legacy eMeasures (Box 11) →Moderate (Moderate is the highest possible rating)

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

Committee pre-evaluation comments

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

2a1. Reliability Specifications

*I don't understand the denominator.the data elements are clearly defined. hard to know if the calculations can be performed consistently since I do not understand the denominator.
the reliability test results of the eMeasure should be comparable to the paper based measure (#2079)"

*The data elements are clear and test results from simulated data set demonstrates measure logic can be interpreted precisely and unambiguously.

*All data elements are clearly defined.
The sample size is small but adequate (• The patient's bundle demographics were designed to mimic the HIV/AIDS population, specifically drawing from the patient characteristicds collected via the Ryan White HIV/AIDS Program Services Report (RSR).)

2a2. Reliability Testing

*The test sample is adequate to generalize for widespread implementation

The results from the Bonnie tool demonstrate sufficient validity so that conclusions can be made about quality
I agree that the results of the eMeasure will be comparable to the chart-abstracted measure (#2079)"

*Reliability was tested with adequate scope using an appropriate method and comparing reliability test results from the paper-based measure NQF#2079 found them to be comparable.

*Is the test sample adequate to generalize for widespread implementation?
Sample is small but adequate.
o Do the results from the Bonnie tool demonstrate sufficient reliability so that differences in performance can be identified?
Yes
o Do you agree that the reliability test results of the eMeasure will be comparable to the paper based measure (#2079)?
Yes.
"

2b1. Validity Specifications
*I don't think there is evidence to support the validity of this measure. Just guidelines.

*No inconsistencies are identified.

*None

2b2. Validity Testing
*Based on the Bonnie testing tool, all test cases performed as expected and the eCQM logic maintained alignment with the clinical intent of the NQF #2079 measure.

*The testing results from the Bonnie tool reached 100% coverage and confirmed there was a test case for each pathway of logic (negative and positive test cases).
The measure had a 100% passing rate which confirmed that all the test cases performed as expected. **"**

2b3-7 Threats to Validity
*2b.3 Exclusions are logical but no evidence presented to see if they are adequate
2b.5 I agree the e-Measure will demonstrate similar results to the chart-abstracted measure.
2b.7 no missing data"

*The eCQM measure shows similar results to the chart-abstracted measure NQF#2079.

* Per the developer, “The HQMF standard specifies that if data are unknown or missing, they shall fail the criterion. This constraint embodies the notion that absence of evidence is evidence of absence, i.e. data not present in a structured field from which the measure draws will not be considered for measure calculation. In certain cases, missing data may have no impact on the measure outcome for a given patient. For example, a data element used in a series of OR statements will not impact the measure outcome if another data element in the OR statement is present and meets all other defined constraints.”
•All Bonnie synthetic patients with missing data performed according to the HQMF standard specification and as expected.

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 provided information on feasibility testing in the eMeasure Feasibility Score Card. The developer did not identify the EHRs used for feasibility testing. Instead, the developer stated that the feasibility assessment was "conducted by consensus of a panel of MITRE clinical informatics, measure development, and eCQM standards experts".
- The developer provided a summary of the latest publicly available data on Meaningful Use EHR capabilities and provider performance on objectives and measures related to the eCQM's data elements:
    - CPOE – Meds
    - CPOE – Labs
    - Demographics
    - Problem List
    - Lab test results
- On a scale from 1 to 3 where 3 is the highest score, all but 3 of the data elements received a score of '3'.
    - Both 'Encounter, Performed: Face to Face' and 'Patient Characteristic Payer' scored a 2 on Data Standards.
        - The Score 2 definition for Data Standards is "terminology standards for this data element are currently available, but it is not consistently coded to standard terminology in the EHR, or the EHR does not easily allow such coding."
    - The data element 'Patient Characteristic Expired' scored a 2 on Data Accuracy. Data accuracy looks at the correctness of the information contained in the data element and whether the data source and recorder are specified. This data element is an exclusion of the measure.
        - The Score 2 definition for Data Acurracy is "the information may not be from the most authoritative source and/or has a moderate likelihood of being correct". The scorecard notes that this information is similar to "self-reporting of a vaccination".
        - The developer notes that "The accuracy of this data element is dependent on full end-to-end interoperability accross providers and between providers and public health agencies."
- The developer indicates that on a scale from 0 to 100 percent, the measure is currently 98.21% feasible and in one to two years, will be 98.81% feasible.
- The measure specifications contain limited proprietary codes for convenience. Users of CPT(R) should obtain all necessary licenses from the owners of these code sets.
- The use of SNOMED Clinical Terms(R) requires a Unified Medical Language System (UMLS) license. These licenses are freely available, from the National Library of Medicine.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*
- *The data element 'Patient Characteristic Expired', the exclusion for this measure, was scored 2 out of 3 for data accuracy on the feasibility scorecard. Does the Committee believe this score impacts the measure's feasibility?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

*The required data elements routinely were routinely generated and used during care delivery at these sites. Further testing will be needed to see how other EHRs work.

*The developer did not identify the EHRs used for feasibility testing, thus the possibility that some EHRs might not be able to routinely generate the data elements, has not been discarded. While this may not be a considerable problem, it would be helpful to assess.

*All data elements are routinely generated.
No issues or concerns.
o Are the required data elements routinely generated and used during care delivery? Yes.
o Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites? Yes, very high.
o The data element 'Patient Characteristic Expired', the exclusion for this measure, was scored 2 out of 3 for data accuracy on the feasibility scorecard. Does the Committee believe this score impacts the measure's feasibility? No. This may be updated during the course of follow up of patient's who don't meet the measure.
"

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**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- This newly developed eMeasure is not currently in an accountability program; however it was reviewed by NQF's Measure Applications Partnership (MAP) for consideration in CMS' Merit Based Incentive Payment Program (MIPS).

Improvement results

- The developer reports performance data from the paper based version of the measure that retention in care has improved over time, stating that of 15,000 patients in the HIVRN database, performance increased from 66.7% in 2007-2008 to 72.6% in 2014-2015.

Unexpected findings (positive or negative) during implementation

- The developer reports that the paper based version of this measure has been adopted by CMS, by the Secretary of the Department of Health and Human Services as a core HIV indicator and in other care settings.

Potential harms

- The developer reports no harms in using the measure.

Vetting of the measure

- According to the developer, the measure has been used in national quality improvement campaigns with participants committing to use the measure, report performance scores and to develop quality improvement project based on the measure. Scores and disparity stratification are shared with participants to benchmark performance.
- In the national quality improvement campaign, data were collected and aggregated every other month. Reports included data tables and spark lines are reported on a public website and via national webinars.

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**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*
- *How has the eCQM 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

* Similar comments to NQF#2079.

*How can the performance results be used to further the goal of high-quality, efficient healthcare?
Follow up of persons who do not meet the measure is possible to identify barriers to care and re-engagement in HIV care.
o Do the benefits of the measure outweigh any potential unintended consequences?
Yes, may help to improve retention in care and viral suppression. No specific unintended consequences.
o How has the eCQM been vetted in real-world settings by those being measure or others?
Limited data provided by developer. Local data available.
"

Criterion 5: Related and Competing Measures

Related or competing measures

- he following measures are listed as related or competing:
  - 2080 Gap in HIV Medical Visits – population but different measurement periods
  - 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 measures listed above. For these measures, the target population is the same (i.e., people living with HIV) however the measure focus is different.

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 candidate 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 x No**

RATIONALE IF NOT ELIGIBLE:

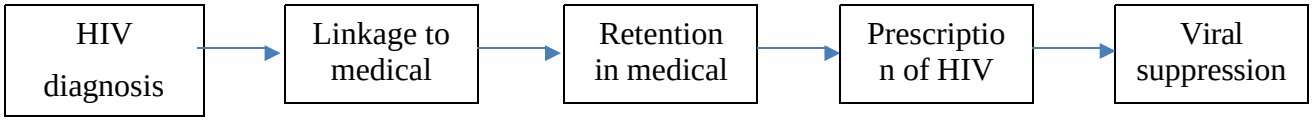
This measure is not eligible for Endorsement + designation since the measure score was tested by face validity only.

Pre-meeting public and member comments

- Please review my comment on NQF#2079 related to the rigidity of frequency measures and their inability to apply to all people with HIV given established practice and clinical guidelines

Measure Title: HIV Medical Visit Frequency

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	**I:** One or more randomized trials with clinical outcomes and/or validated laboratory endpoints
B: Moderate recommendation for the statement	**II:** One or more well-designed, non-randomized trials or observational cohort studies with long-term clinical outcomes
C: Optional recommendation for the statement	**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-636179032321042047.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.)*
There is no measure-specific web page for the electronic version of this measure.

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 an eMeasure **Attachment:**
NQFXXX_MedicalVisitFrequency_Artifacts.zip,NQFXXX_MedicalVisitFrequency_MeasureSubmissionForm-636179038006883388.docx

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:** HIVMVF_v4_6_Thu_Dec_15_20.35.34_CST_2016.xls

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).
Patients who had at least one medical visit in each 6-month of a consecutive consecutive 24 month 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.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).
HIV medical visits are represented by a QDM variable that is comprised of the below seven different encounter type QDM elements:

- Encounter, Performed: Face-to-Face Interaction using Face-to-Face Interaction Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1048)
- Encounter, Performed: Office Visit using Office Visit Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1001)
- Encounter, Performed: Outpatient Consultation using Outpatient Consultation Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1008)
- Encounter, Performed: Preventive Care - Established Office Visit, 0 to 17 using Preventive Care - Established Office Visit, 0 to 17 Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1024)
- Encounter, Performed: Preventive Care Services - Established Office Visit, 18 and Up using Preventive Care Services - Established Office Visit, 18 and Up Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1025)
- Encounter, Performed: Preventive Care Services-Initial Office Visit, 18 and Up using Preventive Care Services-Initial Office Visit, 18 and Up Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1023)
- Encounter, Performed: Preventive Care- Initial Office Visit, 0 to 17 using Preventive Care- Initial Office Visit, 0 to 17 Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1022)

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients, regardless of age, diagnosed with HIV during the first 3 months of the year preceding the measurement period or prior to the measurement period with at least one medical visit in the first 6 months of the year preceding the measurement period.

The target population for this measure is all people living with HIV.

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).
The patient's HIV diagnosis is represented by the QDM element "Diagnosis: HIV" using "HIV Grouping Value Set (2.16.840.1.113883.3.464.1003.120.12.1003)". The patient's medical visits are represented by the following QDM elements:

- Encounter, Performed: Face-to-Face Interaction using Face-to-Face Interaction Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1048)
- Encounter, Performed: Office Visit using Office Visit Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1001)
- Encounter, Performed: Outpatient Consultation using Outpatient Consultation Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1008)
- Encounter, Performed: Preventive Care - Established Office Visit, 0 to 17 using Preventive Care - Established Office Visit, 0 to 17 Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1024)
- Encounter, Performed: Preventive Care Services - Established Office Visit, 18 and Up using Preventive Care Services - Established Office Visit, 18 and Up Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1025)
- Encounter, Performed: Preventive Care Services-Initial Office Visit, 18 and Up using Preventive Care Services-Initial Office Visit, 18 and Up Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1023)
- Encounter, Performed: Preventive Care- Initial Office Visit, 0 to 17 using Preventive Care- Initial Office Visit, 0 to 17 Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1022)

The target population is identified by selecting patients based on their diagnosis with HIV.

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients who died at any time during the measurement period or the 12 months preceding the 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.)*
Denominator exclusions are a subset of the denominator that should not be considered for inclusion in the numerator. This measure denominator exclusion excludes patients who died at any time during the measurement period or the 12 months preceding the measurement period.
Patient death is identified by using the QDM datatype of "Patient Characteristic Expired." In alignment with the CMS/ONC Electronic Clinical Quality Measure Logic and Implementation Guidance Version 1.12 and the Quality Data Model, Version 4.2 and Version 4.3, the "Patient Characteristic Expired" data element is fixed to SNOMED-CT code 41909909 (Dead) and therefore cannot be further qualified with a value set.

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.
Electronic Health Record (Only)

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.
Data is obtained from structured data fields in electronic health 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-636177547706980737.docx,NQFXXX_MedicalVisitFrequency_BonnieTestingAttachment-636177547707136738.zip

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*)**:** Click here to enter NQF number

Measure Title: HIV Medical Visit Frequency
Date of Submission: 12/16/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation

of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Synthetic Bonnie test patients	☒ other: Synthetic Bonnie test patients

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*).

This measure is a legacy electronic clinical quality measure (eCQM) – an NQF endorsed measure that has been respecified into eMeasures and are currently used in federal quality programs. Per NQF modified testing requirements for legacy eCQMs, the measure was tested in the Bonnie testing tool. Bonnie is designed to validate eCQM specifications (HQMF output and value sets) against the measure's expected behavior for user-developed synthetic test patients.

The synthetic patient bundle used to test this measure was designed to simulate clinically relevant, realistic patient scenarios aligned with the target population for this measure. Full details on the Bonnie synthetic patient bundle used to test this measure are included in the Bonnie testing attachment.

For more information on Bonnie, please visit https://bonnie.healthit.gov/.

1.3. What are the dates of the data used in testing? The Bonnie test environment simulates the year 2012 as the measurement period.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

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

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*)

Not applicable. The Bonnie synthetic patient bundle was used to test the measure.

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*)

A test bundle of 64 patients was designed and built within the Bonnie testing tool to evaluate the measure logic. Information documented for each patient within the bundle include:

- Patient name
- Date of birth
- Race
- Ethnicity
- Gender
- Payer

Additional elements contained within the patient profiles as appropriate for testing against expected outcomes include:

- Diagnosis
- Encounters

The patient bundle's demographics were designed to mimic the HIV/AIDS population, specifically drawing from the patient characteristics collected via the Ryan White HIV/AIDS Program Services Report (RSR).

The breakdown of test bundle demographics for the 64 patients included (represented by number of patients/percentage of bundle): males 46/73%; females 17/27%%; American Indian/Alaska Native 2/3
%; Asian 1/2%; Black/African American 30/48%; Native Hawaiian/Pacific Islander 0/0%; White 17/27%; Hispanic/Latino 14/22%; younger than 13 2/3%; 13-17 years old 1/2%; 18-24 years old 2/3%; 25-34 years old 10/16%; 35-44 years old 15/24%; 45-54 years old 21/33%; 55-65 years old 10/16%; older than 65 3/5%.
Full details on the Bonnie synthetic patient bundle used to test this measure, including human-readable and QRDA Category 1 format documents for each synthetic patient record, are included in the Bonnie testing attachment.

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.

The Bonnie patient test deck was used to satisfy all testing requirements for this measure. The testing results are further supported by testing data for the chart-abstracted version of this measure collected through the Health Resources and Services Administration HIV/AIDs Bureau's Ryan White HIV/AIDS Program Services Report.

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).

Patient sociodemographic variables considered in the analysis of the chart-abstracted version of this measure were included in the eCQM specifications and modeled in the Bonnie patient bundle. These variables included age, race, ethnicity, gender and payer.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

Currently, there is no performance data available to test the eCQM. However, the chart-abstracted version of this measure has been in use in national quality reporting programs since as early as 2010.

The most recent reliability analysis of the chart-abstracted measure was confirmed 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.

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*)

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)

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

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)*

The Bonnie testing environment was used to test the validity of the measure logic and data elements. For each Bonnie synthetic patient, an expected measure result was assigned to reflect the expected outcome of the measure given the specific patient scenario and associated data. The synthetic patients were run against the HQMF output loaded into Bonnie, which produces a measure outcome for each patient and evaluates it against the expected outcome. A patient is considered to pass Bonnie testing when the expected outcome matches the actual outcome, e.g. when a patient is expected to be in the numerator population and the computation of the synthetic patient data against the eCQM logic places the patient in the numerator.

In order to achieve a rigorous, clinically relevant test bundle, synthetic patients were designed following the below principles and test areas:

- Clinical relevance. References cited within the chart abstracted measure specification were used to design clinically relevant, realistic patient profiles for the measure's target population. This approach ensured the eCQM logic maintained alignment with the clinical intent of the chart abstracted measure.
- 100% logic coverage: The resulting bundle of synthetic patients collectively includes all data elements and conditions logic that are specified within the measure logic, including at least one patient evaluating against each measure population pathway. Fully testing the measure logic increases test rigor and mitigates risk of unexpected outcomes.
- Edge case testing. Edge cases refer to those data elements that test the upper or lower boundary of measure logic conditions, e.g. a diagnosis starting on the latest qualifying date or medical visits that were exactly 60 days apart. Edge cases are designed to test each edge that exists within each measure population.
- Negative testing. Negative testing involves use of test cases do not evaluate positively against measure logic, but are otherwise clinically relevant and realistic, e.g. scenarios where an HIV diagnosis was not documented or where medical visits did not take place within the expected six month period. Negative testing further validates measure logic by accurately evaluating patients against expected outcomes and simulating the effect of missing data on measure results.

In addition to Bonnie testing, the measure specifications were reviewed independently by three eCQM experts to confirm the logic was syntactically correct, using appropriate and current versions of the eCQM standards and terminologies, and consistent with the intent of the chart-abstracted measure.

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

Bonnie testing results provide logic coverage and passing rates. The synthetic bundle reached 100% coverage, confirming each logic pathway was tested. The results also showed 100% passing rate, confirming all synthetic patients performed as expected.

Full details on Bonnie testing results are contained in the Bonnie testing attachment. The attachment includes a human-readable (HTML) summary document that lists each patient within the bundle and its passing status against expected measure outcomes. The attachment also includes a summary spreadsheet for the synthetic patient bundle which lists each patient, associated demographics, expected and actual measure population outcomes, and which portions or each measure population logic the patient meets expectations for.

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?*)

The results of measure logic testing through use of Bonnie provided confidence in the measure logic accurately representing the clinical intent and alignment with the chart abstracted measure.

2b3. EXCLUSIONS ANALYSIS (FOR MEASURS WITH EXCLUSIONS --- gap in visits and medical visit frequency)
NA ☐ no exclusions — *skip to section 2b4*

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 exclusion – patient death during the measurement period. The exclusion was tested similarly to other criteria using synthetic patients in Bonnie. When the exclusion element was present, the patients were correctly excluded from the measure. In the absence of the exclusion element, cases were not excluded from the 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 that a significant number of patients would be excluded from the measure.

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*)

Exclusions were tested using Bonnie. See response to question 2b.3.1 above.

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*)

Exclusions were tested using Bonnie. See response to question 2b.3.1 above.

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 ___.

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

2b4.1.1 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.

Not applicable.

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.

Not applicable.

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*)

Not applicable.

2b4.4a. What were the statistical results of the analyses used to select risk factors? Not applicable.

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)

Not applicable.

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

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

Not applicable.

If stratified, skip to _____

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:** Not applicable.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*): Not applicable.

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves: Not applicable.

2b4.9. Results of Risk Stratification Analysis: Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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)*

The chart-abstracted version of this measure has been in use since 2010. 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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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*)

The HQMF standard specifies that if data are unknown or missing, they shall fail the criterion. This constraint embodies the notion that absence of evidence is evidence of absence, i.e. data not present in a structured field from which the measure draws will not be considered for measure calculation. In certain cases, missing data may have no impact on the measure outcome for a given patient. For example, a data element used in a series of OR statements will not impact the measure outcome if another data element in the OR statement is present and meets all other defined constraints.

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*)

The Bonnie synthetic patient bundle includes scenarios for missing data elements, which are a form of negative testing. All Bonnie synthetic patients with missing data performed according to the HQMF standard specification and as expected.

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*)

Please see response for question 2b7.1 above.

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: NQFXXX_MedicalVisitFrequency_Feasibility_Scorecard_v1.0-636177547712128770.xlsx

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.
Not applicable.

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)*.
The measure specifications contain limited proprietary codes for convenience. Users of CPT(R) should obtain all necessary licenses from the owners of these code sets.
The use of SNOMED Clinical Terms(R) requires a Unified Medical Language System (UMLS) license. These licenses are freely available, from the National Library of Medicine.

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	
Payment Program	
Quality Improvement (external benchmarking to organizations)	
Quality Improvement (Internal to the specific organization)	

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

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

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

NATIONAL QUALITY FORUM

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 #: 2080
Measure Title: Gap in HIV medical visits
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 did not have a medical visit in the last 6 months of the measurement year

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.
1b.1. Developer Rationale: While prompt linkage to, and sustained retention in, HIV medical care have been clearly shown to maximize patient outcomes, defining and measuring "optimal retention" is not necessarily straightforward, as the most appropriate or useful measure varies according to where the patient is in his/her treatment trajectory (newly diagnosed, recently reengaged in care after some lapse in treatment, or long-time care recipients), who will use the measure (e.g., providers, administrators, or payors), and how the information yielded by the measure will be used.

Retention appears to play a critical role in assisting people living with HIV in their pursuit of achieving viral control and reducing new infections. From an analysis performed by CDC from 2011 data, about 70% of people living with HIV did not have their virus under control. Among the nearly 840,000 people who had not achieved viral suppression, 66% had been diagnosed with HIV, but were not engaged in regular HIV care. A 2015 study estimated the number of HIV transmission from people engaged at the five stages of the HIV care continuum. Ninety-one percent of new HIV infections in 2009 were attributable to people with HIV who were not in medical care, including those who didn't know they were infected. In comparison, less than 6% of new infections could be attributed to people with HIV who were in care and receiving antiretroviral therapy.

It is envisioned that this measure will have a significant impact on patient retention because the patients listed in the numerator are those who require a medical visit. In other words, no additional work needs to be done to generate a list of patients in need of follow-up. A list of the patients in the numerator can be generated, and the medical provider staff can immediately begin follow-up with the patient to schedule an appointment for a medical visit.

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 did not have a medical visit in the last 6 months of the measurement year (Measurement year is a consecutive 12-month period of time).
Denominator Statement: Number of patients, regardless of age, with a diagnosis of HIV who had at least one medical visit in the first 6 months of the measurement year. (The measurement year can be any consecutive 12-month period.)
Denominator Exclusions: Patients who died at any time during the measurement year.

Measure Type: Process
Data Source: Other, Paper Records
Level of Analysis: Clinician : Group/Practice, 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. 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 other guidelines on HIV care and treatment with varying levels of evidence.

Questions for the Committee:

- *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 gaps in medical visits without empirical evidence?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure that monitoring gaps in 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 gaps or 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 Ryan White HIV/AIDS Program Services Report (RSR). The RSR is HRSA HAB's primary source of annual provider and client level data collected from over 2000 recipients and sub-recipients.
- The mean performance for gaps in medical visits has fluctuated over time but stands at 21.7% as of calendar year 2014.

	2014 N=316,087	2013 N=327,618	2012 N=335,408	2011 N=327,744	2010 N=324,455
# of Providers	813	823	816	811	846
Mean	21.7	19.0	17.3	18.8	18.6
Median	15.6	14.3	14.2	14.3	14
10th Percentile	6.5	6.4	6.1	6.3	6.5
90th Percentile	45.1	35.7	31.4	35.9	34.8

Disparities

- The developer provided the following 2010 – 2014 data for gaps in medical visits. The table below shows disparities in gaps in medical visits among patients aged 20-34, among Native Hawaiian/Pacific Islander and American Indian/Alaska Natives, and transgender patients. Numbers are presented as percentages.

Age	2014	2013	2012	2011	2010
<13	12.3	17.4	19.3	21.1	19.3
13-14	9.9	10.8	9.8	16.5	10.4
15-19	14.1	12.8	14.1	14.3	12.3
20-24	21.4	18.4	19.1	18.5	18.1
25-29	21.7	19.6	18.7	18.9	19.0
30-34	20.5	18.9	17.8	17.9	17.4
35-39	19.1	17.9	17.0	16.8	16.4
40-44	17.6	16.7	15.7	16.0	15.4
45-49	16.8	16.1	15.0	14.7	14.8
50-54	15.7	15.6	13.8	13.9	13.9
60-64	14.0	14.2	12.2	12.4	12.3
≥65	13.2	15.2	12.3	13.4	14.0
Race/Ethnicity					
American Indian/Alaska Native	19.9	19.8	15.6	17.2	17.3
Asian	18.2	15.9	13.5	12.8	14.1
Black/African American	17.5	17.3	16.2	16.6	16.0
Hispanic/Latino	15.3	14.2	13.3	13.4	14.6
Native Hawaiian/Pacific Islander	21.2	21.0	18.1	14.4	18.8
White	18.7	17.7	16.0	15.3	14.6
Multiple Races	16.5	14.6	13.7	13.7	12.4

Gender					
Male	18.0	17.1	15.7	15.5	15.5
Female	15.5	15.3	14.7	15.5	14.8
Transgender	19.8	18.8	15.8	16.4	17.9

Questions for the Committee:

- *Does the Committee agree that there is a gap in performance on gaps in medical visit frequency that warrants a national performance measure for continued endorsement?*

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)

1a. Evidence

*"Not an outcome measure.
There is clear evidence linking viral suppression to patient outcomes. It is not clear from the submission that evidence links patient ""retention"" to viral load control, though this approach is logical and supported by guidelines. Also, not clear how to measure ""retention.""
Providers bare some responsibility for assuring retention of their patients."

*I felt the process does a good job of presenting and capturing supporting data. I also think the evaluation process for evaluating the measuring tool helps maintain its validity.

*Process measure with no systematic review available.
• Does the committee agree that viral suppression is a related heath outcome performance measure?
Yes, gaps in medical 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 gaps in medical visits 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 monitoring gaps in medical visit frequency is linked to improved outcomes?
Yes, based on guideline review panel recommendations

1b. Performance Gap
***There is a clear and significant gap between optimal performance of HIV suppression and data from CDC 2014.

*It demonstrated by raw numbers and percentile the amount of people engaged in care and with suppressed Viral Loads versus those who aren't Virally suppressed and not on ARV. Yes the data did a great job of identifying the subgroups based on several sociodemographic factors.i.e sex, age and race

***Does the Committee agree that there is a gap in performance on gaps in medical visit frequency that warrants a national performance measure for continued endorsement?**
Yes, based on the data provided (moderate). The gaps in performance among transgender patients is concerning.

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 did not have a medical visit in the last 6 months of the measurement year. The measurement year is a consecutive 12-month period of time.
- The denominator includes number of patients with HIV regardless of age, who had at least one medical visit in the first 6 months of the measurement year. Patients are excluded if they died at any time during the measurement period.
- The measure calculates a rate where a lower 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:

- Reliability testing was completed for the 2010 – 2014-time period.

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 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).
- The dataset included 2014 HRSA HIV/AIDS Bureau data from the Ryan White HIV/AIDS Program (RWHAP) Services Report (RSR). Over 800 RWHAP sites were included in testing. The RSR is inclusive of the overall RWHAP client population and key priority populations served by RWHAP.

Results of reliability testing

- Testing results showed a median reliability of 0.973 in 2014, which the developer reported demonstrates good reliability.

	2014	2013	2012	2011	2010
Median	0.973	0.970	0.955	0.965	0.969
Minimum	0.147	0.120	0.132	0.105	0.112
Maximum	1.0	1.0	1.0	1.0	1.0

- The developer also gave a distribution of provider-level reliability scores by year

Year	N	>0.9 n (%)	>0.8 n(%)	>0.7 n (%)
2014	813	641 (78.8)	736 (90.5)	764 (94.0)
2013	823	651 (79.1)	731 (88.8)	766 (93.1)
2012	816	583 (71.5)	699 (85.7)	746 (91.4)
2011	811	632 (77.9)	715 (88.2)	756 (93.2)
2010	846	651 (77.0)	742 (87.7)	777 (91.8)

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:

- *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 SC agreed that the measure met the scientific acceptability criteria. Face validity was used to establish measure validity.

Describe any updates to validity testing:

- Validity testing has not been updated.

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 agreed that the measure "could assess and improve the quality of HIV care".

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *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
- To examine the effect of exclusion on the performance score, the developer calculated the proportion of patients excluded due to death out of the total number of patients. The percentage point difference between performance scores with and without the exclusion for death was calculated.
- The developer reports less than 1 percent of patients were excluded due to death each year and had a minimal impact on performance scores for gap in medical care.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. 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, data from 2010-2014 show variability across providers allowing for the identification of meaningful differences across sites. Of the top performers on this measure in 2014, 6.5% of the patients had a gap in care, compared to 45% of patients with a gap in care among the lower performers.

	2014 N=316,087	2013 N=327,618	2012 N=335,408	2011 N=327,744	2010 N=324,455
# of Providers	813	823	816	811	846
Mean	21.7	19.0	17.3	18.8	18.6
Median	15.6	14.3	14.2	14.3	14
10th Percentile	6.5	6.4	6.1	6.3	6.5
90th Percentile	45.1	35.7	31.4	35.9	34.8

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 is not applicable for this measure based on the method used to calculate gaps in medical care. Specifically, the logic used to determine the number of patients with a gap in medical care relied on whether or not the patient had at least one medical visit in the first six months of the measurement year, and then among these patients, whether or not the patient had at least one medical visit during the last 180 days of that year.

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

*The data elements are all clearly specified, the calculation logic is clear and it is likely the measure can be consistently applied. Consultants who only see the patient intermittently may be disadvantaged.

*Performance Measure score. No additional step are required. I have no concerns as the measure seems to a good job of predicting in its structure any potential changes that might influence any barriers to its implementation..

*"o Are all the data elements clearly defined? Are all appropriate codes included? Yes, reliability is rated as high.
o Is the logic or calculation algorithm clear? Yes
o Is it likely this measure can be consistently implemented? Yes, used consistently at the local level.
"

2a2. Reliability Testing

*No need for retesting reliability.

*I thought the reliability was adequate base on the number of organizations and patients involved. There was no established Cut off for minimum reliability level.

*o 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?
No. Data provided is adequate.
If the Committee does not choose to re-vote, then a discussion may still be needed.
o Is the measure score test sample adequate to generalize for widespread implementation?
o Do the results demonstrate sufficient reliability so that differences in performance can be identified?
"

2b1. Validity Specifications

**"There is clear evidence linking viral suppression to patient outcomes. It is not clear from the submission that evidence links patient ""retention"" and viral load control, though this approach is logical and supported by guidelines.
Providers bare some responsibility for assuring retention of their patients."

*I felt the evidence supported the PRO.

*Specification are consistent with the evidence.

2b2 Validity Testing
*There is clear evidence linking viral suppression to patient outcomes. It is not clear from the submission that evidence links patient "retention" and viral load control, though this approach is logical and supported by guidelines. It is not clear that a visit in the last 6 months of the measurement period is the correct measure, though it is logical and appropriate to clinical norms of care.

**Systematic assessment of face validity of performance measure score as an indicator. I found the results demostrated sufficient validity in it conclusions based on the statistical analysis done by technical work group.

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

2b3-7. Threats to Validity
*"2b.3 Should patient ""leaving the practice"" be considered in exclusions?
2b.5 the measure identifies meaningful differences in care.
2b.7 It's hard to know how complete patient appointment data is. How are missed/cancelled appointments counted?"

*I found no threats to validity. They excluded patients who passed away during the research so they impact measurements. I thought the results clearly showed a link between early entry in to care as indicator of adherence and outcomes.

*Are the exclusions consistent with the evidence?
Yes, death only can be identified.
o Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?
Death only represents <1% of patients excluded.

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 all data 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.
- There are no fees, licensing, or other requirements to use this 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

***"This is a paper-derived measure.
It's hard to know how complete patient appointment data is. How are missed/cancelled appointments counted?"

*A full analysis of missing data is routinely done. Using defined fields all data elements can be captured. The only concerns I would have about the data collection is the system being used to collect it.

* Are the required data elements routinely generated and used during care delivery?
Yes.
o 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

- The measure is not currently used in the 20+ federal programs under the Centers for Medicare & Medicaid Services.
- This measure is used in the Ryan White HIV/AIDS program:
  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

Improvement results

- The developer reports that retention in care has become less of a priority even though a significant number of HIV transmissions have been linked to people who are not retained in care. The RWHAP has experienced a 3-point increase in gap in medical visits from 18.6% in 2010 to 21.7% in 2014.

Unexpected findings (positive or negative) during implementation

- The developer reports that adoption and use of this measure has continued to spread since it was developed and has been submitted for adoption to CMS measurement programs. This measure was selected as one of the core HIV indicators by the Secretary of the Department of Health and Human Services.

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 has been received from Ryan White HIV/AIDS Program grant recipients and subrecipients regarding the feasibility and usefulness of the data presented in the Ryan White HIV/AIDS Program Annual Client-Level Data Report Grant recipient report using the data for benchmarking their program, setting goals/targets, and gaining a fuller understanding of all aspects of the Ryan White HIV/AIDS Program (i.e. other regions of the country). Grant recipients and subrecipients have also requested additional analyses.

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 from RWHAP recipients 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**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*
- *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 widely used across Ryan White programs, but it appears that it has not led to improved care (as defined by the measure).
What is meant by, ""RWHAP grant recipients have encouraged further stratification, dissemination methods, and graphical presentations?""
What is meant by, ""Additional feedback from RWHAP recipients 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?""
Is this measure better than or needed in addition to 2079 or 2082?
Will it be supplanted by 3209?

*I think its been vetted and has a system in place to do so on an ongoing basis. I think the usability is broad in the area of QA in terms of Accountability and performance.

*"o How can the performance results be used to further the goal of high-quality, efficient healthcare?
Providers can follow up with persons who have a gap in care to re-engage them in care.
o Do the benefits of the measure outweigh any potential unintended consequences?
Yes, a small number of patients who are not out of care may be contacted to return to care.
o How has the measure been vetted in real-world settings by those being measure or others?
Yes, used locally to identify persons lost to care. "

Criterion 5: Related and Competing Measures

Related or competing measures

- 2079 HIV Medical Visit Frequency
- 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)
- 3209 HIV Medical Visit Frequency (newly submitted eMeasure)
- 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 measures listed above. For these measures, the target population is the same (i.e., people living with HIV) however the measure focus is different.

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 candidate 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:

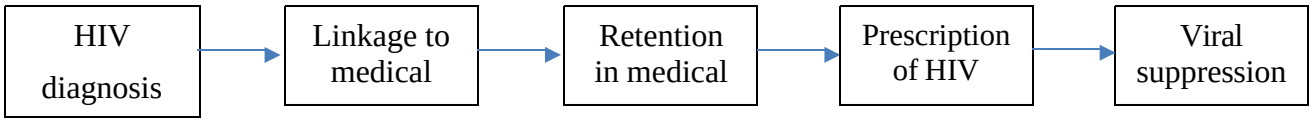
This measure is not eligible for Endorsement + designation since the developer did not perform empirical validity testing of the measure score.

Pre-meeting public and member comments

- Gaps in care should be focused on minimum standards applying to all patients as absence of a clinical visit within a 12 month period.

Measure Title: Gap in HIV Medical Visits

1a.12 LOGIC MODEL



Although the above diagram outlines the sequential steps 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	**I:** One or more randomized trials with clinical outcomes and/or validated laboratory endpoints
B: Moderate recommendation for the statement	**II:** One or more well-designed, non-randomized trials or observational cohort studies with long-term clinical outcomes
C: Optional recommendation for the statement	**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

Gap_evidence.docx,Gap_submission.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.

While prompt linkage to, and sustained retention in, HIV medical care have been clearly shown to maximize patient outcomes, defining and measuring "optimal retention" is not necessarily straightforward, as the most appropriate or useful measure varies according to where the patient is in his/her treatment trajectory (newly diagnosed, recently reengaged in care after some lapse in treatment, or long-time care recipients), who will use the measure (e.g., providers, administrators, or payors), and how the information yielded by the measure will be used.

Retention appears to play a critical role in assisting people living with HIV in their pursuit of achieving viral control and reducing new infections. From an analysis performed by CDC from 2011 data, about 70% of people living with HIV did not have their virus under control. Among the nearly 840,000 people who had not achieved viral suppression, 66% had been diagnosed with HIV, but were not engaged in regular HIV care. A 2015 study estimated the number of HIV transmission from people engaged at the five stages of the HIV care continuum. Ninety-one percent of new HIV infections in 2009 were attributable to people with HIV who were not in medical care, including those who didn't know they were infected. In comparison, less than 6% of new infections could be attributed to people with HIV who were in care and receiving antiretroviral therapy.

It is envisioned that this measure will have a significant impact on patient retention because the patients listed in the numerator are those who require a medical visit. In other words, no additional work needs to be done to generate a list of patients in need of follow-up. A list of the patients in the numerator can be generated, and the medical provider staff can immediately begin follow-up with the patient to schedule an appointment for a medical visit.

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 (<u>current and over time</u>) at the specified level of analysis**. *(<u>This is required for maintenance of endorsement</u>. 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 "Gap submission" 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.**
See 1b2.

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.** *(<u>This is required for maintenance of endorsement</u>. 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 "Gap submission" 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.)*

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:** Gap_data_dictionary.docx

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.
N/A

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 did not have a medical visit in the last 6 months of the measurement year (Measurement year is a consecutive 12-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 not have had a medical visit in the last 6 months of the measurement year.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Number of patients, regardless of age, with a diagnosis of HIV who had at least one medical visit in the first 6 months of the measurement year. (The measurement year can be any consecutive 12-month 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 year
2. Patients without a date of death during the measurement year
3. Patients diagnosed with HIV during the first 3 months of the measurement year or prior to the measurement year
4. Patients who had at least one medical visit in the first 6 months of the measurement year

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

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 year.

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 = Lower 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.) had a HIV diagnosis prior to the measurement year or during the first three months of the measurement year; 2.) did not have a date of death during the measurement year; and 3.) had at least one medical visit in the first 6 months of the measurement year. 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: did not have a medical visit in the last 6 months of the measurement year.
3. Calculate the percentage 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.

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Other, 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.
Not applicable.

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)*
Clinician : Group/Practice, 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.)*
N/A

2. Validity – See attached Measure Testing Submission Form
Gap_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.)
Yes

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*)**:** 2080
Measure Title: Gap In HIV Medical Visits
Date of Submission: Click here to enter a date
Type of Measure:

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

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured;* On an annual basis, Ryan White HIV/AIDS Program (RWHAP) grant recipient and subrecipients submit the Ryan White HIV/AIDS Services Report (RSR). The RSR dataset is the Health Resources and Services Administration HIV/AIDS Bureau's primary source of annual, client-level data collected from its nearly 2,000 funded grant recipients and subrecipients. Since 2010, client-level RSR data have been used to assess the numbers and types of clients receiving services and their HIV outcomes. Project Officers at the HIV/AIDS Bureau share the data with grant recipients and subrecipients to monitor and support their progress at improving care and treatment for people living with HIV. It is through the hard work of these providers and the RWHAP community that clients are helped every day.

RSR includes all clients served by the RWHAP during calendar years 2010 through 2014. RSR data do not include information about AIDS Drug Assistance Programs (ADAP); all ADAP-related information is collected through another data system. Although data presented in this report are "nonADAP," this does not imply the clients did not receive ADAP services. ADAP data will be published separately, at later time.

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: Click here to describe	☐ other: Click here to describe

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

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. Over 800 (varies by year) Ryan White HIV/AIDS Program outpatient ambulatory medical care providers representing various types, locations, and sizes were included in the testing.

Descriptive characteristics of RWHAP providers

	2010		**2011**		**2012**		**2013**		**2014**	
	N	%	N	%	N	%	N	%	N	%
Overall	846		811		816		823	--	813	--
Provider type										
Hospital or university-based clinic	355	17.5	358	18.6	349	19.1	351	19.6	338	19.4
Community based organization	1,114	54.9	1,053	54.8	993	54.3	958	53.6	921	53.0
Health department	284	14.0	274	14.3	243	13.3	233	13.0	243	14.0
Other	275	13.6	237	12.3	243	13.3	247	13.8	237	13.6
HHS Region										

Region 1	149	8.0	153	8.6	142	8.4	139	8.4	135	8.3
Region 2	368	19.7	339	19.0	323	19.1	303	18.3	293	18.1
Region 3	180	9.6	177	9.9	174	10.3	174	10.5	160	9.9
Region 4	337	18.0	335	18.8	312	18.5	301	18.1	313	19.3
Region 5	197	10.5	189	10.6	177	10.5	188	11.3	180	11.1
Region 6	150	8.0	142	8.0	133	7.9	131	7.9	132	8.2
Region 7	65	3.5	60	3.4	57	3.4	56	3.4	54	3.3
Region 8	48	2.6	43	2.4	34	2.0	35	2.1	46	2.8
Region 9	300	16.0	281	15.7	277	16.4	276	16.6	253	15.6
Region 10	78	4.2	68	3.8	60	3.6	56	3.4	52	3.2

1.6. How many and which <u>patients</u> 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.

Distribution of patients per provider by year, 2010-2014

Year	N patients, mean	N patients, median	Min patients	Max patients
2010	384	177	1	13,159
2011	404	182	1	13,380
2012	411	179	1	13,849
2013	398	181	1	14,755
2014	388	177	1	13,850

Descriptive characteristics of RWHAP patients by year, 2010-2014

	2010		2011		2012		2013		2014	
	No.	%	No.	%	No.	%	No.	%	No.	%
OVERALL	324,455	–	327,744	–	335,408	–	327,618	–	316,087	–
AGE GROUP										
<13	3,709	1.2	3,647	1.1	3,150	1.0	2,667	0.9	2,720	0.9
13–14	627	0.2	605	0.2	469	0.1	360	0.1	343	0.1
15–19	3,698	1.2	3,541	1.1	3,066	0.9	2,609	0.8	2,506	0.8
20–24	14,040	4.5	14,831	4.6	15,741	4.8	15,538	5.0	14,578	4.8
25–29	22,120	7.0	23,278	7.3	24,904	7.7	25,586	8.2	26,043	8.5
30–34	28,644	9.1	29,330	9.2	30,084	9.3	29,495	9.4	28,484	9.3
35–39	35,161	11.2	33,597	10.5	33,005	10.2	31,560	10.1	30,691	10.0
40–44	50,769	16.1	47,941	15.0	45,343	14.0	40,728	13.0	37,000	12.1
45–49	60,344	19.2	59,453	18.6	58,145	17.9	52,863	16.8	47,932	15.6
50–54	46,433	14.7	48,647	15.2	50,876	15.7	50,491	16.1	50,492	16.4
55–59	28,015	8.9	30,646	9.6	33,215	10.2	33,493	10.7	34,667	11.3
60–64	13,441	4.3	15,237	4.8	16,991	5.2	17,780	5.7	19,399	6.3
≥65	8,187	2.6	8,946	2.8	10,147	3.1	10,780	3.4	12,231	4.0
RACE/ETHNICITY										
American Indian/ Alaska Native	1,473	0.5	1,366	0.4	1,371	0.4	1,414	0.5	1,272	0.4
Asian	3,382	1.1	3,598	1.2	3,980	1.2	3,835	1.2	3,791	1.2
Black/ African American	146,460	47.3	149,834	47.8	150,974	47.2	146,056	47.0	142,746	46.9
Hispanic/Latino[a]	71,002	22.9	71,240	22.7	75,201	23.5	74,967	24.1	74,714	24.5
Native Hawaiian/ Pacific Islander	627	0.2	710	0.2	575	0.2	510	0.2	442	0.2
White	83,854	27.1	83,061	26.5	83,820	26.2	78,953	25.4	75,931	24.9
Multiple races	3,177	1.0	3,716	1.2	4,238	1.3	4,899	1.6	5,651	1.9
GENDER										
Male	219,625	69.7	223,379	69.9	230,075	70.8	221,930	70.7	216,965	70.7
Female	93,266	29.6	93,687	29.3	92,186	28.4	89,212	28.4	87,071	28.4
Transgender	2,313	0.7	2,585	0.8	2,848	0.9	2,779	0.9	2,974	1.0
TRANSMISSION RISK										
Male client										
Male-to-male sexual contact	117,267	59.9	120,622	60.2	128,744	61.8	127,571	62.2	127,624	62.7
Injection drug use	17,479	8.9	16,787	8.4	15,586	7.5	15,509	7.6	13,753	6.8
Male-to-male sexual contact and injection drug use	6,971	3.6	6,837	3.4	6,974	3.3	6,136	3.0	6,396	3.1
Heterosexual contact	48,903	25.0	50,814	25.4	52,266	25.1	51,174	24.9	51,155	25.1
Perinatal infection	3,830	2.0	3,919	2.0	3,604	1.7	3,419	1.7	3,456	1.7
Other	1,248	0.6	1,231	0.6	1,309	0.6	1,402	0.7	1,189	0.6
Female client										
Injection drug use	9,264	11.2	9,022	10.7	8,182	9.8	8,310	10.0	7,396	9.1
Heterosexual contact	68,009	82.4	69,767	82.8	70,362	84.1	69,356	83.9	69,090	84.8
Perinatal infection	4,338	5.3	4,587	5.4	4,182	5.0	4,003	4.8	4,093	5.0
Other	900	1.1	877	1.0	936	1.1	1,044	1.3	940	1.2

Transgender client										
Sexual contact	1,874	90.7	2,058	91.2	2,281	91.8	2,314	92.9	2,499	93.2
Injection drug use	38	1.8	32	1.4	35	1.4	32	1.3	31	1.2
Sexual contact and injection drug use	144	7.0	156	6.9	158	6.4	130	5.2	135	5.0
Perinatal infection	5	0.2	5	0.2	2	0.1	4	0.2	9	0.3
Other	6	0.3	5	0.2	8	0.3	10	0.4	8	0.3
HEALTH CARE COVERAGE										
Private only	35,392	12.4	37,532	12.3	39,972	12.7	37,204	12.1	–	–
Medicare only	23,245	8.1	24,279	8.0	23,538	7.5	22,840	7.5	–	–
Medicaid only	73,292	25.6	75,690	24.8	71,990	22.8	69,211	22.6	–	–
Other public	22,398	7.8	20,977	6.9	28,039	8.9	27,347	8.9	–	–
Other private	11,512	4.0	9,884	3.2	6,049	1.9	3,682	1.2	–	–
No coverage	86,220	30.1	100,001	32.8	103,150	32.7	101,524	33.1	–	–
Multiple coverages	34,276	12.0	36,330	11.9	42,969	13.6	44,578	14.6	–	–
Private employer	–	–	–	–	–	–	–	–	18,805	6.3
Private individual	–	–	–	–	–	–	–	–	16,154	5.4
Medicare	–	–	–	–	–	–	–	–	26,145	8.7
Medicaid	–	–	–	–	–	–	–	–	94,993	31.6
Medicare and Medicaid	–	–	–	–	–	–	–	–	19,207	6.4
Veterans Administration	–	–	–	–	–	–	–	–	454	0.2
Indian Health Service	–	–	–	–	–	–	–	–	71	0.0
Other plan	–	–	–	–	–	–	–	–	11,899	4.0
No coverage	–	–	–	–	–	–	–	–	90,828	30.2
Multiple coverages	–	–	–	–	–	–	–	–	22,428	7.5

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.

Ryan White HIV/AIDS Program Services Report (RSR) 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 (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

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.

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*)

Provider-level reliability results for the "gap in medical care" measure in 2014 are detailed below. Results for years 2010-2013 are available upon request, but were not included due to space constraints.

Provider-level "gap in medical care" reliability testing (signal to noise) results, 2014.

Site/provider ID	% patients w/ gap	variance within	reliability
55	48.8%	0.001	0.973
63	26.1%	0.008	0.720
82	44.6%	0.003	0.889
88	22.0%	0.000	0.992
96	13.9%	0.000	0.991
101	24.6%	0.001	0.941
105	16.0%	0.001	0.959
112	7.1%	0.000	0.985
113	9.9%	0.000	0.983
117	7.2%	0.000	0.997
118	9.5%	0.000	0.988

120	20.9%	0.001	0.952
123	9.0%	0.000	0.992
124	16.7%	0.001	0.970
127	18.9%	0.001	0.976
128	27.7%	0.000	0.981
133	25.6%	0.000	0.996
135	16.1%	0.001	0.956
138	6.8%	0.001	0.937
140	18.2%	0.014	0.615
141	20.5%	0.001	0.966
143	6.5%	0.000	0.978
144	21.0%	0.001	0.978
147	6.6%	0.000	0.979
148	40.8%	0.005	0.814
149	21.7%	0.001	0.936
154	16.9%	0.000	0.989
155	4.3%	0.002	0.923
156	19.6%	0.003	0.863
158	28.4%	0.002	0.934
159	8.9%	0.000	0.983
160	20.2%	0.002	0.923
164	28.6%	0.003	0.881
168	18.4%	0.000	0.991
169	11.3%	0.001	0.968
170	9.0%	0.000	0.985
171	12.5%	0.000	0.982
172	10.8%	0.000	0.985
173	11.0%	0.001	0.971
174	6.5%	0.000	0.992
175	14.2%	0.001	0.965
176	12.0%	0.000	0.997
177	9.5%	0.001	0.940
178	12.5%	0.000	0.997
179	3.7%	0.000	0.980
181	13.4%	0.000	0.990
182	22.6%	0.000	0.991
183	5.7%	0.000	0.995
184	3.8%	0.000	0.995
186	2.0%	0.000	0.991

187	9.6%	0.000	0.990
188	37.3%	0.001	0.977
191	5.6%	0.000	0.993
192	42.5%	0.000	0.980
194	9.7%	0.000	0.982
196	10.0%	0.000	0.996
197	12.9%	0.000	0.995
199	20.7%	0.000	0.985
201	9.7%	0.000	0.999
203	6.3%	0.000	0.997
205	58.3%	0.020	0.516
207	4.0%	0.000	0.995
209	41.1%	0.001	0.945
210	95.8%	0.000	0.985
211	12.6%	0.000	0.989
212	8.2%	0.000	0.985
213	17.7%	0.000	0.995
214	6.9%	0.000	0.996
215	12.2%	0.000	0.986
216	4.9%	0.000	0.996
217	13.8%	0.000	0.978
220	2.7%	0.000	0.996
221	19.8%	0.001	0.961
222	7.7%	0.000	0.991
223	11.8%	0.000	0.997
224	9.8%	0.000	0.992
225	12.5%	0.000	0.997
227	11.3%	0.000	0.995
228	8.1%	0.000	0.982
230	10.9%	0.000	0.998
231	17.0%	0.000	0.990
232	29.3%	0.001	0.960
233	10.8%	0.000	0.996
235	7.7%	0.001	0.977
236	12.6%	0.000	0.986
238	7.2%	0.000	0.997
239	16.6%	0.000	0.991
240	21.4%	0.000	0.982
241	17.2%	0.000	0.998

242	13.0%	0.001	0.971
244	6.2%	0.000	0.994
245	7.1%	0.000	0.995
246	12.3%	0.000	0.996
248	15.4%	0.000	0.997
252	9.1%	0.000	0.988
253	16.8%	0.000	0.992
255	17.3%	0.001	0.937
256	23.8%	0.001	0.973
257	7.9%	0.000	0.993
259	16.2%	0.001	0.946
263	80.4%	0.000	0.988
265	21.3%	0.000	0.997
266	14.8%	0.000	0.994
267	13.1%	0.000	0.995
268	16.7%	0.000	0.980
269	6.7%	0.000	0.994
271	26.4%	0.003	0.889
273	15.6%	0.000	0.996
275	28.3%	0.000	0.991
276	17.4%	0.001	0.976
277	26.4%	0.000	0.990
278	33.6%	0.000	0.986
279	10.6%	0.000	0.996
280	78.1%	0.005	0.802
283	30.2%	0.000	0.991
284	29.9%	0.001	0.967
285	23.6%	0.001	0.966
286	6.5%	0.000	0.993
288	15.5%	0.001	0.959
289	37.9%	0.001	0.966
290	41.6%	0.003	0.873
291	22.1%	0.000	0.988
292	12.0%	0.000	0.978
294	16.9%	0.000	0.995
295	11.6%	0.000	0.985
298	14.0%	0.000	0.996
299	16.1%	0.000	0.991
302	18.0%	0.000	0.991

303	12.5%	0.000	0.983
304	14.1%	0.000	0.993
305	29.8%	0.000	0.991
307	17.0%	0.000	0.993
308	6.5%	0.001	0.974
310	15.0%	0.000	0.998
311	39.3%	0.001	0.949
312	37.0%	0.003	0.895
313	20.0%	0.001	0.953
314	13.5%	0.000	0.995
315	13.8%	0.000	0.995
316	19.0%	0.001	0.971
317	17.4%	0.001	0.969
318	15.5%	0.000	0.987
319	13.8%	0.001	0.959
320	12.5%	0.001	0.969
321	5.8%	0.000	0.982
322	16.7%	0.001	0.962
323	10.1%	0.000	0.988
324	21.0%	0.001	0.967
325	10.4%	0.000	0.998
326	10.8%	0.000	0.994
328	17.5%	0.000	0.988
329	11.4%	0.000	0.999
332	11.5%	0.000	0.989
333	22.7%	0.002	0.923
334	14.4%	0.000	0.983
335	14.1%	0.000	0.989
336	7.9%	0.000	0.992
340	7.9%	0.000	0.989
342	9.8%	0.000	0.989
343	11.6%	0.000	0.979
344	13.9%	0.000	0.991
345	11.4%	0.001	0.961
347	8.6%	0.000	0.997
348	11.0%	0.000	0.986
349	14.8%	0.000	0.982
351	9.8%	0.000	0.993
353	11.7%	0.000	0.993

357	12.9%	0.000	0.995
358	72.5%	0.001	0.939
360	10.6%	0.000	0.994
361	13.8%	0.000	0.991
362	11.8%	0.002	0.934
363	5.2%	0.000	0.981
365	2.4%	0.001	0.975
366	9.6%	0.000	0.978
368	7.3%	0.000	0.991
369	20.3%	0.001	0.969
370	4.5%	0.000	0.978
371	12.5%	0.001	0.954
372	13.6%	0.000	0.995
375	10.3%	0.000	0.998
378	6.2%	0.001	0.961
379	14.6%	0.000	0.995
380	18.8%	0.000	0.986
382	13.0%	0.000	0.992
384	12.0%	0.000	0.981
385	12.1%	0.001	0.977
386	19.5%	0.001	0.960
388	14.1%	0.000	0.996
389	6.9%	0.000	0.985
390	5.8%	0.000	0.993
391	22.6%	0.001	0.964
393	16.5%	0.000	0.992
394	16.1%	0.002	0.908
395	11.7%	0.000	0.993
400	7.0%	0.000	0.987
404	7.8%	0.000	0.979
407	26.2%	0.001	0.976
408	6.1%	0.000	0.996
409	16.0%	0.001	0.953
410	13.9%	0.002	0.934
412	13.5%	0.000	0.988
414	10.9%	0.001	0.966
417	14.3%	0.001	0.952
421	22.1%	0.000	0.988
422	7.6%	0.000	0.992

423	8.7%	0.000	0.990
425	13.2%	0.000	0.978
427	17.0%	0.000	0.996
438	15.7%	0.000	0.989
441	8.0%	0.001	0.957
457	9.0%	0.000	0.997
460	.	.	.
463	17.4%	0.000	0.979
469	15.5%	0.000	0.982
473	0.0%	0.000	1.000
480	24.6%	0.001	0.977
481	40.8%	0.001	0.963
483	16.9%	0.000	0.992
489	9.5%	0.000	0.994
491	24.7%	0.002	0.912
498	7.7%	0.000	0.994
504	11.5%	0.000	0.997
506	56.1%	0.001	0.965
509	17.3%	0.000	0.991
510	24.2%	0.001	0.936
517	6.1%	0.000	0.991
534	3.4%	0.000	0.997
553	52.2%	0.011	0.666
593	23.0%	0.001	0.937
598	16.8%	0.001	0.974
612	6.8%	0.001	0.968
664	6.5%	0.000	0.991
704	12.4%	0.001	0.976
710	22.0%	0.003	0.863
726	32.4%	0.001	0.967
738	27.8%	0.011	0.660
744	12.2%	0.000	0.988
753	36.4%	0.001	0.966
757	40.2%	0.002	0.922
762	63.2%	0.000	0.983
765	41.6%	0.001	0.957
775	32.1%	0.002	0.913
783	62.2%	0.005	0.805
787	33.3%	0.001	0.937

791	47.3%	0.001	0.966
793	4.8%	0.002	0.909
794	65.0%	0.004	0.851
798	36.8%	0.000	0.986
799	31.5%	0.004	0.844
800	25.1%	0.001	0.956
801	16.7%	0.005	0.824
803	17.4%	0.000	0.982
807	29.7%	0.001	0.960
818	11.9%	0.000	0.996
820	26.6%	0.002	0.912
821	37.8%	0.003	0.883
824	25.0%	0.004	0.835
841	43.0%	0.003	0.891
852	63.8%	0.002	0.928
861	42.8%	0.001	0.953
867	14.1%	0.000	0.994
871	20.1%	0.000	0.994
873	18.0%	0.001	0.976
894	12.6%	0.000	0.997
905	18.5%	0.000	0.992
907	32.1%	0.000	0.996
913	17.7%	0.000	0.995
920	6.7%	0.004	0.839
926	18.9%	0.004	0.839
927	12.4%	0.001	0.957
929	7.2%	0.000	0.978
933	7.4%	0.000	0.985
945	10.9%	0.001	0.964
980	5.1%	0.000	0.984
986	7.9%	0.001	0.977
992	3.7%	0.000	0.999
996	51.7%	0.000	0.987
1009	5.9%	0.001	0.975
1017	4.4%	0.000	0.979
1022	20.7%	0.001	0.970
1023	2.6%	0.000	0.994
1026	23.8%	0.002	0.926
1029	35.2%	0.001	0.938

1031	8.7%	0.000	0.993
1036	15.0%	0.000	0.985
1037	39.4%	0.002	0.904
1038	13.9%	0.000	0.983
1049	24.1%	0.000	0.980
1050	15.4%	0.000	0.993
1052	18.5%	0.006	0.795
1055	26.5%	0.001	0.974
1056	7.6%	0.000	0.996
1066	10.7%	0.000	0.992
1067	17.3%	0.000	0.990
1068	19.4%	0.000	0.995
1093	19.1%	0.002	0.905
1094	10.0%	0.000	0.995
1100	9.5%	0.000	0.993
1109	14.9%	0.000	0.994
1110	8.4%	0.001	0.971
1112	13.5%	0.000	0.997
1120	9.6%	0.000	0.995
1121	12.5%	0.000	0.997
1122	10.1%	0.000	0.990
1131	11.2%	0.000	0.995
1132	2.3%	0.000	0.992
1146	0.0%	0.000	1.000
1155	24.0%	0.000	0.981
1160	12.3%	0.000	0.996
1162	13.0%	0.002	0.912
1163	46.5%	0.001	0.963
1167	16.9%	0.000	0.998
1214	7.5%	0.000	0.979
1216	8.3%	0.001	0.968
1229	5.9%	0.000	0.991
1230	9.6%	0.001	0.973
1263	100.0%	0.000	1.000
1276	16.2%	0.000	0.980
1278	42.7%	0.002	0.901
1284	8.2%	0.001	0.965
1287	6.5%	0.000	0.997
1289	8.1%	0.002	0.915

1300	45.2%	0.004	0.844
1302	70.3%	0.001	0.972
1309	6.0%	0.001	0.950
1310	28.6%	0.004	0.856
1314	66.0%	0.000	0.981
1318	15.1%	0.002	0.925
1319	18.9%	0.002	0.930
1333	19.6%	0.001	0.975
1349	21.1%	0.003	0.881
1358	15.6%	0.001	0.960
1359	11.8%	0.000	0.998
1364	14.1%	0.002	0.927
1378	13.5%	0.001	0.963
1380	43.6%	0.003	0.873
1382	13.4%	0.001	0.948
1401	45.3%	0.000	0.981
1430	25.0%	0.005	0.806
1444	24.3%	0.005	0.813
1445	13.7%	0.000	0.997
1448	0.0%	0.000	1.000
1451	22.6%	0.002	0.920
1456	17.8%	0.000	0.996
1461	15.1%	0.001	0.955
1464	10.0%	0.000	0.979
1479	25.0%	0.016	0.580
1490	95.0%	0.001	0.948
1511	24.8%	0.000	0.987
1512	17.1%	0.001	0.965
1514	32.6%	0.005	0.809
1527	65.0%	0.011	0.655
1552	12.3%	0.000	0.997
1567	46.7%	0.003	0.887
1570	56.2%	0.002	0.931
1572	40.0%	0.005	0.818
1574	19.8%	0.002	0.929
1582	33.8%	0.001	0.939
1583	25.0%	0.047	0.316
1587	16.7%	0.012	0.651
1594	22.0%	0.003	0.881

1597	7.4%	0.003	0.895
1607	20.7%	0.001	0.968
1610	12.6%	0.001	0.977
1628	56.3%	0.008	0.738
1634	36.5%	0.002	0.906
1635	56.9%	0.002	0.906
1637	38.0%	0.001	0.966
1650	15.2%	0.001	0.954
1654	17.6%	0.003	0.884
1656	9.9%	0.001	0.969
1668	36.9%	0.000	0.989
1672	51.9%	0.009	0.700
1684	13.3%	0.000	0.997
1719	15.2%	0.000	0.995
1762	13.3%	0.000	0.994
1784	11.5%	0.001	0.971
1786	56.9%	0.004	0.851
1792	44.8%	0.000	0.987
1806	12.5%	0.014	0.613
1809	11.7%	0.000	0.994
1812	13.1%	0.000	0.993
1831	4.1%	0.001	0.976
1834	2.4%	0.001	0.974
1847	12.5%	0.001	0.950
1849	57.7%	0.002	0.926
1879	42.9%	0.001	0.944
1900	0.0%	0.000	1.000
1904	14.3%	0.004	0.832
1912	48.2%	0.000	0.982
1930	19.3%	0.001	0.938
1955	23.5%	0.002	0.933
1967	57.6%	0.007	0.745
1968	6.7%	0.002	0.912
1970	87.5%	0.014	0.613
1972	100.0%	0.000	1.000
1977	0.0%	0.000	1.000
1980	14.6%	0.003	0.893
1989	18.5%	0.001	0.939
2003	75.0%	0.047	0.316

2008	16.7%	0.023	0.483
2010	30.0%	0.021	0.507
2011	47.6%	0.012	0.645
2017	18.2%	0.005	0.827
2020	90.0%	0.009	0.706
2025	11.6%	0.000	0.983
2028	11.5%	0.001	0.976
2029	10.3%	0.001	0.953
2034	18.9%	0.003	0.882
2041	62.5%	0.029	0.425
2049	30.0%	0.005	0.805
2058	22.0%	0.001	0.973
2072	45.5%	0.008	0.742
2073	13.4%	0.000	0.993
2076	39.1%	0.010	0.676
2078	34.4%	0.007	0.754
2080	33.3%	0.074	0.226
2081	0.0%	0.000	1.000
2116	27.8%	0.002	0.921
2117	8.7%	0.003	0.862
2118	18.5%	0.002	0.930
2126	20.9%	0.001	0.946
2127	34.4%	0.003	0.896
2129	5.7%	0.000	0.983
2133	20.7%	0.000	0.979
2134	32.3%	0.007	0.754
2137	15.4%	0.005	0.812
2139	95.1%	0.000	0.987
2141	9.3%	0.000	0.989
2143	13.7%	0.000	0.994
2148	21.8%	0.001	0.940
2150	10.2%	0.000	0.997
2153	10.8%	0.001	0.972
2163	12.7%	0.001	0.975
2170	19.0%	0.000	0.978
2174	8.9%	0.000	0.999
2175	7.4%	0.000	0.983
2178	12.4%	0.000	0.988
2180	11.8%	0.000	0.995

2183	6.9%	0.000	0.983
2187	34.8%	0.010	0.687
2188	38.9%	0.013	0.621
2189	60.0%	0.012	0.643
2191	36.7%	0.004	0.848
2198	42.9%	0.005	0.812
2200	27.3%	0.006	0.782
2203	52.8%	0.002	0.904
2205	84.9%	0.002	0.899
2207	36.8%	0.000	0.995
2224	10.3%	0.000	0.994
2228	33.3%	0.012	0.636
2230	52.2%	0.002	0.907
2232	0.0%	0.000	1.000
2246	13.3%	0.001	0.978
2252	25.4%	0.003	0.878
2263	4.0%	0.002	0.934
2264	50.0%	0.125	0.147
2296	20.3%	0.001	0.976
2299	29.6%	0.000	0.982
2320	10.3%	0.000	0.996
2366	17.4%	0.002	0.912
2368	42.6%	0.004	0.857
2374	100.0%	0.000	1.000
2378	10.5%	0.005	0.813
2379	77.3%	0.008	0.730
2381	11.7%	0.001	0.970
2388	0.0%	0.000	1.000
2389	100.0%	0.000	1.000
2415	23.3%	0.001	0.940
2420	29.0%	0.002	0.929
2436	12.2%	0.001	0.973
2438	8.6%	0.001	0.941
2444	0.0%	0.000	1.000
2457	5.6%	0.000	0.989
2474	21.4%	0.001	0.971
2495	34.3%	0.003	0.865
2514	11.1%	0.000	0.999
2525	12.2%	0.000	0.994

2572	13.9%	0.000	0.991
2654	65.8%	0.006	0.785
2694	0.0%	0.000	1.000
2699	44.4%	0.014	0.612
2700	9.4%	0.000	0.991
2702	15.2%	0.000	0.986
2703	4.9%	0.000	0.996
2704	9.9%	0.000	0.988
2707	0.0%	0.000	1.000
2709	12.5%	0.000	0.995
2714	31.0%	0.001	0.962
2717	15.3%	0.001	0.954
2718	11.5%	0.000	0.998
2720	36.3%	0.002	0.921
2721	9.8%	0.000	0.996
2723	14.0%	0.000	0.984
2728	7.3%	0.000	0.988
2732	18.0%	0.000	0.984
2737	33.3%	0.005	0.803
2745	12.6%	0.001	0.959
2746	24.0%	0.001	0.937
2757	20.4%	0.001	0.968
2764	18.6%	0.001	0.963
2766	7.8%	0.000	0.979
2767	27.3%	0.018	0.545
2768	18.5%	0.000	0.996
2769	27.4%	0.003	0.871
2771	89.2%	0.000	0.987
2775	7.9%	0.000	0.996
2779	18.0%	0.001	0.970
2782	16.6%	0.001	0.967
2788	15.0%	0.001	0.972
2790	17.0%	0.000	0.987
2794	59.6%	0.005	0.824
2795	22.0%	0.001	0.943
2849	59.6%	0.001	0.971
2855	10.2%	0.001	0.968
2856	28.3%	0.004	0.849
2857	22.7%	0.008	0.730

2865	34.9%	0.000	0.981
2872	85.0%	0.003	0.871
2873	57.1%	0.035	0.382
2892	11.5%	0.000	0.983
2893	15.8%	0.001	0.977
2928	13.1%	0.000	0.989
2929	10.6%	0.001	0.951
2933	13.6%	0.001	0.970
2945	19.0%	0.001	0.963
2947	21.9%	0.000	0.990
2948	19.5%	0.000	0.984
2949	15.0%	0.001	0.956
2951	15.6%	0.001	0.971
2958	22.2%	0.005	0.818
2959	21.2%	0.003	0.871
2966	16.7%	0.001	0.949
2969	11.4%	0.001	0.937
2970	7.0%	0.001	0.971
2988	1.6%	0.000	0.989
2989	0.0%	0.000	1.000
2997	11.9%	0.000	0.995
3010	35.8%	0.001	0.970
3012	14.4%	0.001	0.970
3020	25.0%	0.047	0.316
3048	42.1%	0.013	0.627
3052	18.0%	0.001	0.951
3077	32.1%	0.003	0.886
3079	9.0%	0.001	0.973
3111	65.2%	0.003	0.868
3131	14.3%	0.001	0.974
3133	15.7%	0.001	0.936
3164	8.2%	0.002	0.934
3177	7.3%	0.000	0.981
3187	80.8%	0.002	0.911
3210	13.1%	0.000	0.983
3255	21.9%	0.002	0.902
3261	6.6%	0.000	0.981
3262	14.0%	0.003	0.886
3264	17.2%	0.005	0.815

3265	21.1%	0.001	0.958
3310	37.1%	0.001	0.948
3359	13.6%	0.001	0.958
3389	100.0%	0.000	1.000
3397	47.0%	0.001	0.953
3401	6.5%	0.001	0.942
3428	8.8%	0.000	0.992
3433	9.4%	0.000	0.981
3440	8.0%	0.000	0.987
3444	12.6%	0.001	0.976
3445	48.2%	0.004	0.829
3449	25.0%	0.004	0.835
3456	7.9%	0.000	0.987
3469	33.1%	0.001	0.963
3482	7.5%	0.001	0.971
3484	12.0%	0.000	0.986
3498	10.0%	0.005	0.828
3507	27.3%	0.004	0.857
3519	16.2%	0.000	0.982
3521	17.5%	0.002	0.923
3523	9.1%	0.001	0.969
3527	6.4%	0.000	0.989
3530	18.3%	0.001	0.966
3533	66.7%	0.025	0.467
3538	7.4%	0.001	0.971
3551	36.4%	0.001	0.967
3552	22.7%	0.008	0.730
3553	15.4%	0.003	0.896
3554	29.6%	0.008	0.737
3573	26.0%	0.001	0.942
3579	12.0%	0.001	0.967
3583	11.7%	0.000	0.996
3587	15.0%	0.002	0.931
3591	16.2%	0.004	0.855
3593	6.7%	0.000	0.988
3594	10.4%	0.001	0.967
3595	6.5%	0.000	0.996
3596	1.8%	0.000	0.993
3597	11.5%	0.000	0.981

3598	100.0%	0.000	1.000
3600	35.0%	0.006	0.792
3601	14.1%	0.000	0.993
3602	9.4%	0.001	0.956
3603	7.3%	0.001	0.975
3604	8.3%	0.006	0.772
3605	16.8%	0.001	0.957
3606	51.6%	0.001	0.943
3607	11.5%	0.000	0.991
3608	28.6%	0.001	0.961
3609	28.0%	0.004	0.843
3625	14.7%	0.002	0.921
3626	42.1%	0.004	0.835
3633	13.6%	0.001	0.937
3639	9.8%	0.000	0.996
3658	90.0%	0.009	0.706
3659	64.5%	0.002	0.898
3687	40.6%	0.002	0.928
3702	9.7%	0.000	0.995
3728	0.0%	0.000	1.000
3769	7.8%	0.000	0.995
3794	30.2%	0.001	0.967
3826	14.6%	0.001	0.943
3847	24.7%	0.000	0.990
3862	72.7%	0.018	0.545
3879	18.7%	0.001	0.971
3904	10.9%	0.000	0.995
3932	50.0%	0.003	0.881
3942	89.3%	0.001	0.967
3959	33.8%	0.003	0.873
3969	12.7%	0.002	0.915
3972	7.2%	0.000	0.983
3973	7.9%	0.000	0.987
3975	37.1%	0.007	0.764
3976	11.2%	0.000	0.988
3977	30.0%	0.000	0.986
3978	21.8%	0.000	0.985
3979	14.0%	0.000	0.989
3980	12.3%	0.000	0.981

3981	10.4%	0.000	0.998
3982	7.7%	0.000	0.992
3983	25.2%	0.001	0.971
3984	37.0%	0.004	0.833
3985	48.9%	0.005	0.803
3998	26.2%	0.002	0.920
4014	7.6%	0.001	0.976
4018	24.3%	0.002	0.926
4035	15.2%	0.001	0.959
4039	37.1%	0.002	0.907
4068	55.0%	0.012	0.636
4088	19.5%	0.001	0.968
4122	11.1%	0.001	0.947
4126	11.4%	0.001	0.937
4220	23.1%	0.001	0.941
4221	42.1%	0.003	0.871
4235	9.6%	0.001	0.948
7685	23.6%	0.001	0.947
7718	60.0%	0.003	0.878
7722	51.4%	0.003	0.865
7728	30.2%	0.004	0.845
7734	4.7%	0.000	0.992
7738	14.1%	0.000	0.994
7758	0.0%	0.000	1.000
7765	12.6%	0.000	0.987
7766	9.4%	0.000	0.992
7772	22.1%	0.000	0.982
7780	66.7%	0.037	0.368
7782	50.0%	0.014	0.609
7785	7.2%	0.000	0.989
7802	15.2%	0.000	0.989
7807	31.1%	0.001	0.937
7821	26.8%	0.003	0.887
7833	26.7%	0.001	0.937
7834	0.0%	0.000	1.000
7845	37.4%	0.001	0.953
7857	28.6%	0.029	0.426
7878	30.4%	0.005	0.824
7885	14.0%	0.000	0.987

7886	20.4%	0.001	0.938
7888	18.2%	0.000	0.983
7892	21.3%	0.001	0.951
7910	0.0%	0.000	1.000
7911	10.7%	0.003	0.863
7913	16.9%	0.002	0.916
7919	24.1%	0.006	0.774
7920	9.6%	0.001	0.948
7929	9.6%	0.000	0.993
7931	12.4%	0.000	0.988
7942	18.7%	0.000	0.990
7955	35.9%	0.002	0.925
7964	13.9%	0.003	0.867
7985	22.6%	0.003	0.885
7997	6.9%	0.000	0.987
7998	23.5%	0.005	0.803
8000	20.6%	0.000	0.991
8005	6.9%	0.000	0.989
8018	23.3%	0.001	0.962
8027	4.7%	0.001	0.969
8029	21.2%	0.001	0.977
8030	17.8%	0.003	0.869
8063	12.2%	0.002	0.908
8067	16.8%	0.000	0.981
8079	12.6%	0.001	0.977
8102	75.7%	0.003	0.892
8111	28.2%	0.002	0.930
8119	58.2%	0.001	0.969
8129	17.7%	0.001	0.971
8130	7.6%	0.000	0.981
8131	9.6%	0.000	0.994
8132	11.6%	0.000	0.993
8133	7.3%	0.000	0.996
8134	1.6%	0.000	0.999
8135	4.7%	0.000	0.995
8136	6.0%	0.000	0.989
8142	15.1%	0.001	0.935
8143	5.6%	0.000	0.978
8149	100.0%	0.000	1.000

8160	13.5%	0.000	0.992
8163	17.9%	0.003	0.892
8166	9.3%	0.001	0.950
8167	14.3%	0.009	0.712
8180	7.7%	0.001	0.965
8181	12.2%	0.001	0.948
8199	39.0%	0.000	0.992
8228	3.7%	0.000	0.995
8229	11.1%	0.001	0.947
8242	15.3%	0.000	0.999
8260	22.0%	0.003	0.881
8261	16.4%	0.001	0.971
8262	31.3%	0.003	0.871
8263	13.9%	0.000	0.989
8265	8.8%	0.000	0.996
8277	15.0%	0.001	0.972
8282	12.5%	0.007	0.760
8284	66.7%	0.009	0.700
8288	0.0%	0.000	1.000
8294	33.3%	0.002	0.917
8295	10.8%	0.001	0.949
8300	50.0%	0.125	0.147
8301	83.3%	0.023	0.483
8302	85.7%	0.017	0.553
8305	0.0%	0.000	1.000
8308	0.0%	0.000	1.000
8313	12.5%	0.000	0.986
8320	0.0%	0.000	1.000
8323	11.5%	0.001	0.963
8330	0.0%	0.000	1.000
8332	37.5%	0.029	0.425
8334	0.0%	0.000	1.000
8355	16.8%	0.001	0.971
8369	37.8%	0.001	0.957
8373	21.4%	0.001	0.964
8387	45.1%	0.001	0.965
8396	54.4%	0.000	0.985
8397	10.3%	0.000	0.988
8399	27.3%	0.005	0.827

8401	3.5%	0.000	0.986
8403	15.3%	0.000	0.986
8405	20.0%	0.002	0.904
8407	20.4%	0.001	0.950
8411	11.8%	0.000	0.995
8412	42.9%	0.035	0.382
8414	13.9%	0.003	0.867
8415	8.3%	0.006	0.772
8419	44.3%	0.001	0.957
8421	17.4%	0.001	0.960
8425	18.6%	0.004	0.860
8426	27.3%	0.001	0.944
8427	14.3%	0.003	0.896
8430	16.7%	0.008	0.737
8432	0.0%	0.000	1.000
8438	29.2%	0.009	0.715
8441	8.0%	0.003	0.880
8504	36.4%	0.021	0.507
8506	27.1%	0.000	0.997
8507	18.4%	0.000	0.979
8508	28.0%	0.001	0.941
8510	14.5%	0.000	0.996
8511	32.7%	0.000	0.986
8512	18.9%	0.000	0.988
8513	25.3%	0.002	0.905
8537	43.1%	0.002	0.906
8538	7.2%	0.000	0.991
8542	31.5%	0.000	0.984
8546	12.5%	0.001	0.957
8550	10.6%	0.001	0.977
8551	10.8%	0.000	0.978
8553	26.7%	0.007	0.768
8559	0.0%	0.000	1.000
8561	11.7%	0.000	0.987
8563	23.9%	0.003	0.894
8566	0.0%	0.000	1.000
8568	25.6%	0.005	0.816
8570	11.1%	0.011	0.663
8571	12.5%	0.001	0.964

8573	25.8%	0.003	0.875
8575	19.1%	0.002	0.929
8577	12.1%	0.003	0.870
8579	33.3%	0.037	0.368
8580	43.4%	0.001	0.955
8598	18.3%	0.000	0.996
8608	19.6%	0.003	0.863
8611	0.0%	0.000	1.000
8618	6.1%	0.000	0.984
8624	18.4%	0.000	0.990
8626	7.7%	0.005	0.798
8632	22.9%	0.001	0.973
8635	.	.	.
8636	.	.	.
8638	.	.	.
8639	.	.	.
8640	.	.	.
8641	.	.	.
8644	.	.	.
8645	.	.	.
8650	16.0%	0.001	0.960
8651	13.8%	0.001	0.972
8653	0.0%	0.000	1.000

Overall reliability scores (i.e., median of provider-level reliability [R_median], minimum [R_min], maximum [R_max]) by year, and the overall variance between sites, are summarized below.

Overall reliability scores by year, 2010-2014

Year	% w/ gap	Var_between	R_median	R_min	R_max
2010	15.3%	0.0158	0.969	0.112	1.000
2011	15.5%	0.0146	0.965	0.105	1.000
2012	15.4%	0.0113	0.955	0.132	1.000
2013	16.6%	0.0170	0.970	0.120	1.000
2014	17.3%	0.0216	0.973	0.147	1.000

Reliability varied across providers by year. The proportion of providers with reliability greater than or equal to 0.9, 0.8, and 0.7 are shown below.

Distribution of provider-level reliability scores by year, 2010-2014

Year	N	≥0.9 n (%)	≥0.8 n (%)	≥0.7 n (%)

2010	846	651 (77.0)	742 (87.7)	777 (91.8)
2011	811	632 (77.9)	715 (88.2)	756 (93.2)
2012	816	583 (71.5)	699 (85.7)	746 (91.4)
2013	823	651 (79.1)	731 (88.8)	766 (93.1)
2014	813	641 (78.8)	736 (90.5)	764 (94.0)

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

There is no established cut-off for minimum reliability level. Values above 0.7 are considered sufficient to see differences between providers and the mean, and values above 0.9 are considered sufficient to see differences between pairs of providers (RAND Corporation, TR-653-NCQA, 2009).

Each year, more than 72%-79% of providers had reliability scores of 0.9 or greater; more than 90% of providers had reliability scores of 0.7 or greater. Therefore, the reliability of viral suppression can be considered to be sufficient to identify real differences in performance across providers. As previously mentioned, sample size is another driver of reliability and likely contributed to lower provider-level reliability (e.g., in 2014 site 2264 had a reliability of 0.147, and reported 1 of 2 patients with a medical visit in the first six months had not had a medical visit in the last 180 days). However, overall (median) reliability was consistently greater than 0.95 during 2010-2014, supporting the conclusion that the reliability of this measure can be considered very good.

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*)

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)*

3. 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

4. Face validity of the performance score was gained through structured presentations (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*)

3. 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.

4. Sixty-nine individuals/organizations submitted 239 pieces of comments. Twenty-three 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?*

3. 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.
4. 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 (23) 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 MEASURES WITH EXCLUSIONS --- gap in visits and medical visit frequency)

NA ☒ no exclusions — *skip to section 2b4*

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*)

To examine the effect of excluding patients due to death during the measurement year on the performance score for the proportion of patients with a gap in medical care, we calculated the proportion of patients excluded due to death out of the total number of patients. The percentage point difference between performance scores with and without the exclusion for death was also calculated, to determine the magnitude of the effect of the exclusion on the performance score.

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*)

Year	% excluded due to death	% age point difference (w/ – w/out exclusion)
2010	0.93	0.88
2011	0.84	0.82
2012	0.81	0.91
2013	0.63	0.74
2014	0.64	0.71

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*)

Less than 1 percent of patients were excluded due to death each year; this exclusion criterion had a minimal impact on performance scores for gap in medical care. In general, the exclusion criteria for death resulted in a very minimal (< 1%) increase in the proportion of patients with a gap in medical care.

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** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

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.

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*)

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

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)

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

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

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

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

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

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

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

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

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

To examine meaningful differences in performance, we examined the distribution of the proportion of patients with a gap in medical care across providers, by year. Performance scores were broken into the bottom 10% and top 90% providers to better characterize the gaps that remain across providers. Following the NHAS 2020 goal of increasing the percentage of persons with HIV infection who are retained in HIV medical care to 90 percent, performance scores were examined with respect the proportion of providers with less than 10% of patients that had a gap in medical care in a given measurement year. The National HIV/AIDS Strategy 2020 sets a goal of 90% for retention; therefore, suggesting a 10% gap in HIV medical visits.

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*)

	% patients with gap in medical care across providers					providers with <10% patients w/ gap		
Year	Mean	SD	Median	10th %ile	90th %ile	N	n	%
2010	18.6	17.1	14.0	6.5	34.8	846	221	26.1
2011	18.8	17.1	14.3	6.3	35.9	811	220	27.1
2012	17.3	13.9	14.2	6.1	31.4	816	212	26.0
2013	19.0	16.7	14.3	6.4	35.7	823	193	23.5
2014	21.7	18.6	15.6	6.5	45.1	813	202	24.8

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. Across years, the tenth percentile of scores indicates that the top performers had roughly 6% of patients with a gap in medical care. In contrast, the 90th percentile of scores can be examined to identify the lowest performers; in 2014, among the bottom 10% of providers, 45% or more of their patients had a gap in medical care. These differences demonstrate the continued value of the measure in identifying sites based on relative performance.

Provider-level performance differences can also underscore important changes in the proportion of patients with a gap in medical care. In 2014, providers had, on average, 20% of patients with a gap in medical care. In particular, among 813 providers, 202 (24.8%) had less than 10% of patients with a gap in medical care. Given the large population that the RWHAP serves, differences in performance across providers – in particular, the lowest performers, represent a substantial number of patients that could benefit from improved retention in care.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

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*)

Based on the method used to calculate the gap in medical care performance score, conducting missing data analysis is not applicable for this measure. Specifically, the logic used to determine the number of patients with a gap in medical care relied on whether or not the patient had at least one medical visit in the first six months of the measurement year, and then among these patients, whether or not the patient had at least one medical visit during the last 180 days of that year. Based on provider reporting, patients were classified as either having a medical visit or not, during both time points, and missing/unknown were not response options.

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*)

N/A (see 2b7.1)

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 (see 2b7.1)

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).
For this measure, we are currently field testing (beta testing) an electronically specified measure.

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 collection and availability: The data used for testing and operational use of this measure are readily available within patient health records and provided annually to the Ryan White HIV/AIDS Program through the reporting of the Ryan White Service Report (approved by the Office of Management and Budget 0915-0323).

Missing data: A full analysis of missing data is provided in this submission.

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 Ryan White HIV/AIDS Program https://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio Payment Program Ryan White HIV/AIDS Program https://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio 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 maintenance of endorsement), 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

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.
The Ryan White HIV/AIDS Program served more than 300,000 unduplicated patients annually between 2010-2014 across 2,000+ grant recipients and subrecipients. Gap in HIV medical visits has seem small increases over the past few years. Antidotal information suggests that retention has become less of a priority even though a significant number of HIV transmission have been linked to people who are not retained in medical care. The Ryan White HIV/AIDS Program has experienced a 3-point increase in gap in HIV medical visits from 18.6% in 2010 to 21.7% in 2014. This increase has been experienced across most of the demographic populations.

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 submitted for adoption or 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.
Starting in 2015, Health Resources and Services Administration began releasing December 1st – World AIDS Day – an annual data report (Ryan White HIV/AIDS Program Annual Client-Level Data Report) that contains data similar to those presenting in the report. Building upon the success of the state profiles (http://hab.hrsa.gov/stateprofiles/), Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived). A supplemental report exploring data for the eligible metropolitan areas and transitional grant areas and youth/young adults has been released as well as slides sets for fact sheets by program and population, special populations (http://hab.hrsa.gov/publications/hivaids-bureau-fact-sheets), and infographics (contained in fact sheets). Additionally, grant recipient level reports are prepared and disseminated to all Ryan White HIV/AIDS Program grant recipients.

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.**
Starting in 2015, Health Resources and Services Administration began releasing December 1st – World AIDS Day – an annual data report (Ryan White HIV/AIDS Program Annual Client-Level Data Report) that contains data similar to those presenting in the report. Building upon the success of the state profiles (http://hab.hrsa.gov/stateprofiles/), Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived). A supplemental report exploring data for the eligible metropolitan areas and transitional grant areas and youth/young adults has been released as well as slides sets for fact sheets by program and population, special populations (http://hab.hrsa.gov/publications/hivaids-bureau-fact-sheets), and infographics (contained in fact sheets). Additionally, grant recipient level reports are prepared and disseminated to all Ryan White HIV/AIDS Program grant recipients.

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 from Ryan White HIV/AIDS Program grant recipients and subrecipients regarding the feasibility and usefulness of the data presented in the Ryan White HIV/AIDS Program Annual Client-Level Data Report. Significant feedback has been provided about the timeliness and expansions of the data release. Grant recipient report using the data for benchmarking their program, setting goals/targets, and gaining a fuller understanding of all aspects of the Ryan White HIV/AIDS Program (i.e. other regions of the country). Grant recipients and subrecipients have also requested additional analyses. Health Resources and Services Administration responded with supplemental reports (Ryan White HIV/AIDS Program Supplemental Client-Level Data Report, Eligible Metropolitan Areas and Transitional Grant Areas; special population reports); slide decks for the overall client population and special populations; grant recipient reports; and infographics – all of which will be updated and released annually. Health Resources and Services Administration plans to release additional analyses and special reports this year based on feedback from Ryan White HIV/AIDS Program grant recipients and subrecipients.

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

Ryan White HIV/AIDS Program national partners (national organizations that represent grant recipients, subrecipients, and patients) has provided antidotal feedback regarding the timeliness, feasibility, and usability of the release of the Ryan White HIV/AIDS Program Annual Client-Level Data Report, supplemental reports, slide decks, fact sheets, and infographics. The national partners encourage the continued release of the data in all its formats.

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

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.

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.**

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)**

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
2079 HIV Medical Visit Frequency
208 HIV Viral Suppression
2083 Prescription of HIV Antiretroviral Therapy

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.**

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.)
N/A

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.

Available at measure-specific web page URL identified in S.1 **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

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.

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:** It is our intention that this measure will be used in quality improvement in addition to public reporting. As it is involved in quality improvement, it is not our intent that the performance goal will be 0%. When we do set the performance goal, we will take into consideration appropriate reasons why the patient may not be able to meet the numerator criterion.

NATIONAL QUALITY FORUM

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 #: 2082
Measure Title: HIV viral load suppression
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 with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year

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: Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. Antiretroviral therapy delays this progression and increases the length of survival.

Antiretroviral therapy 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). Viral suppression is a main goal of HIV treatment and an indicator of treatment success and reduction of potential HIV transmission. It is directly related to:

- Reduction in disease progression, incidence of opportunistic infections, the risk of both defining and non-AIDS- defining complications and the incidence and severity of chronic conditions.
- Reduction in the risk of transmitting HIV to a sexual or drug-using partner who does not have HIV.
- Improvement of immune function, quality of life, increase in time until development of AIDS increase in life expectancy.

Being virally suppressed is good for an HIV-positive person's overall health and preventing HIV infection from advancing to AIDS, the last stage of HIV infection.

- Durable viral suppression improves immune function and quality of life, prolongs life, lowers the risk of both AIDS-defining and non-AIDS-defining complications, and prolongs life. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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 with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year
Denominator Statement: Number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the measurement year
Denominator Exclusions: There are no patient exclusions.

Measure Type: Outcome
Data Source: Laboratory, 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 health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary:

- 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.
- According to the developer, viral suppression is a main goal of HIV treatment and an indicator of treatment success and reduction of potential HIV transmission.
- Being virally suppressed is good for a HIV-positive person's overall health and preventing HIV infection from advancing to AIDS, the last stage of HIV infection. Viral suppression is directly related to:
    - Reduction in disease progression, incidence of opportunistic infections, the risk of both defining and non-AIDS- defining complications and the incidence and severity of chronic conditions.
    - Reduction in the risk of transmitting HIV to a sexual or drug-using partner who does not have HIV.
    - Improvement of immune function, quality of life, increase in time until development of AIDS and increase in life expectancy.
- The developer provided multiple guidelines for the administration of antiretroviral therapy and viral load monitoring intervals for adults, adolescents and pregnant women.
- The 2012 Infectious Disease Steering Committee agreed that there is a substantial relationship between viral load suppression and the reduction of morbidity, mortality and HIV transmission. The 2012 SC also agreed that emerging evidence of earlier antiretroviral therapy indicated decreased HIV-associated complications.

Changes to evidence from last review

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

Updates: See evidence summary above

- The developer provided sufficient evidence demonstrating that antiretroviral therapy and viral suppression reduce morbidity and mortality associated with HIV.

Question for the Committee:

- *Does the Committee agree that a viral load of less than 200 copies/mL leads to improved patient outcomes for patients with a diagnosis of HIV?*
- *Does the Committee agree that there is at least one thing that the provider can do to achieve a change in the viral load of patients diagnosed with HIV? If so, does the Committee agree there is no need for repeat discussion and vote on Evidence?*

Guidance from the Evidence Algorithm: Health outcome measure (Box 1)→The relationship between the outcome and at least one process is identified and supported by the stated rationale (Box 2)→Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

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 provided the following facility-level performance rates from the Ryan White HIV/AIDS Program Services Report (RSR) from 2010 – 2014:

	2014	2013	2012	2011	2010
Rate	80.8	76.1	69.9	65.5	61.8
Pts w/ ≥1 medical visit (den)	316,087	327,618	335,408	327,744	324,455
Pts w/viral suppression (num)	255,342	249,436	234,505	214,650	200,584
Mean	80.3	76.1	69.9	64.7	60.6
Median	84.2	80.7	75.6	71.4	67.8
Standard Deviation	15.5	17.0	20.3	22.1	23.8
10th percentile	65.0	57.1	40.2	31.9	19.5
90th percentile	93.1	90.2	88.0	84.9	82.8
Min, Max	0.0, 100.0	0.0, 100.0	0.0, 100.0	0.0, 100.0	0.0, 100.0
Pts w/viral load test performed	293,237 (92.8)	297,066 (90.7)	289,563 (86.3)	273,241 (83.4)	264,630 (81.6)
# of facilities	813	823	816	811	846

Disparities:

- The developer provided the following 2010 – 2014 viral suppression rates from the measure as specified:

Age	2014	2013	2012	2011	2010
<13	35.5	37.9	36.3	30.5	36.3

13-14	83.8	81.6	76.2	69.2	60.9
15-19	71.5	65.4	57.3	53.8	51.0
20-24	68.2	60.2	50.7	46.9	41.8
25-29	72.5	66.3	58.3	53.5	48.9
30-34	75.9	70.5	63.5	59.5	55.2
35-39	78.0	73.7	67.3	63.0	60.0
40-44	79.9	75.9	70.3	66.0	62.5
45-49	82.1	78.4	72.9	68.5	64.9
50-54	85.7	81.9	77.0	72.0	67.8
60-64	87.1	83.4	78.5	73.8	69.7
≥65	88.4	84.7	80.7	74.7	70.7
Race/Ethnicity					
American Indian/Alaska Native	82.4	74.2	72.3	68.2	64.9
Asian	83.4	78.5	72.4	67.6	64.8
Black/African American	77.4	72.3	66.0	61.2	56.9
Hispanic/Latino	82.6	78.2	72.7	67.6	62.8
Native Hawaiian/Pacific Islander	75.4	67.9	65.2	67.9	57.9
White	83.8	80.2	74.5	70.3	68.3
Multiple Races	83.6	78.0	71.7	66.0	66.8
Gender					
Male	80.7	76.1	70.2	65.7	62.3
Female	79.8	75.0	68.8	63.7	58.9
Transgender	73.6	72.1	66.3	60.1	55.4

- The developer also provided performance rates based on transmission risk, health care coverage, provider type and National HIV/AIDS Strategy (NHAS) populations.

Questions for the Committee:

- *Does the data demonstrate that there continues to be a quality problem and variation in viral suppression for patients diagnosed with HIV? Is there opportunity for improvement?*
- *Is a national performance measure still warranted?*
- *Do you agree the data demonstrates a disparity in care for various populations?*

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)

1a. Evidence

*Evidence is provided that viral load suppression is a marker of improved health outcomes for persons with HIV. Less evidence is provided as to why a level of fewer than 200 copies/ml was chosen as the marker of viral suppression, and not < 50copies/ml or "undetectable".

*"I agree that a viral load of less than 200 copies/mL leads to improved patient outcomes for patients with a diagnosis of HIV
I agree that there is at least one thing that the provider can do to achieve a change in the viral load of patients diagnosed with HIV
I agree there is no need for repeat discussion and vote on Evidence?"

*"This is an outcome measure up for maintenance review. HIV viral load suppression is linked with decreased disease progression, incidence of OIs, and other clinically relevant outcomes.

I am not aware of any new studies that alters the evidence base.

However, the viral load indicated by the viral load is set at <200. With improvements in antiretroviral therapy and assays to measure viral load, guidelines support viral suppression which would now be considered levels much lower than 200 (i.e. less than 20 or undetectable). Current performance data suggests there is still opportunity for improvement even with a permissive cutoff of 200. The developer should consider reassessing the cutoff range in the future as rates of compliance increase. "

*Viral load is related to several health-related outcomes (i.e., disease progression, incidence of opportunistic infections, etc.), all of which appear to be of direct clinical benefit. These outcomes appear to be positive correlated to viral load. I am not aware of any recent studies supporting these outcomes.

1b. Performance Gap

*Evidence is provided demonstrating variability in achieving viral load suppression across providers and with variability in performance in population subgroups. Evidence is provided of improvement over time, but gaps persist.

*"The data demonstrates that there continues to be a quality problem and variation in viral suppression for patients diagnosed with HIV
There is an opportunity for improvement?
A national performance measure is still warranted
The data demonstrates a disparity in care for various populations
"

*Performance data indicates that there is a gap in performance that warrants a national performance measure. The data demonstrates variation in performance by demographic groups. Children <13 in particular show significantly lower rates of treatment success, as measured by viral load. Rates are also lower in younger adults, although not nearly as pronounced as among those <13.

*Yes, there are performance aps related to age and to ethnicity, these have improved from 2010-2014 as given in data. These gaps warrant the measure.

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): Laboratory, Paper Medical Records

Specifications:

- The level of analysis is at the facility-level.
- The numerator includes the number of patients in the denominator with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year.
- The denominator includes the number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the measurement year. To be included in the denominator, patients must meet **all** of the following conditions/events:
  - Patients of any age during the measurement year
  - Patients diagnosed with HIV during the first 3 months of the measurement year or prior to the measurement year
  - Patients who had at least one medical visit during the measurement year
- There are no patient exclusions.
- There are no codes (ICD-9, ICD-10, CPT, etc.) included in the specifications.
- The calculation algorithm is included.

Questions for the Committee:

- *Are all the data elements clearly defined?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure is 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 2013, measure score reliability using a beta-binomial model to assess the signal-to-noise ratio was conducted using 9 of 18 HIV Research Network (HIVRN) sites and 11,134 patients in 2010. The clinic-specific reliability ranged from 0.93 to 0.99 with a between clinic variance of 0.0066.
- The Steering Committee noted that the testing data for reliability was well-defined.

Describe any updates to testing: The developer provided updated measure score reliability testing – see below

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 2010 - 2014 HRSA HIV/AIDS Bureau data from the Ryan White HIV/AIDS Program (RWHAP) Services Report (RSR). The RSR is inclusive of the overall RWHAP client population and key priority populations served by RWHAP. Over 800 Ryan White HIV/AIDS Program providers and more than 315,000 patients (varies by year) were included in the testing.
- The developers used a beta-binomial model to assess the signal-to-noise ratio. A reliability of 0.00 implies that all the variability in a measure is attributable to measurement error. A reliability of 1.0 implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one physician from another. This is an appropriate test for measure score reliability. A reliability of 0.70 is generally considered a minimum threshold for reliability.

Results of reliability testing:

- The developer provided the following overall reliability scores by year from 2010 – 2014:

Year	Variance Between	Median Reliability	Min Reliability	Max Reliability
2010	0.051	0.983	0.290	1.000
2011	0.046	0.982	0.267	1.000
2012	0.038	0.979	0.338	1.000
2013	0.020	0.967	0.211	1.000
2014	0.013	0.954	0.092	1.000

- The developer provided the distribution of provider-level reliability scores by year from 2010 – 2014:

Year	N	≥0.9 n (%)	≥0.8 n (%)	≥0.7 n (%)
2010	846	764 (90.3)	809 (95.6)	826 (97.6)
2011	811	721 (88.9)	766 (94.5)	786 (96.9)
2012	816	713 (87.4)	775 (95.0)	794 (97.3)
2013	823	657 (79.8)	738 (89.7)	772 (93.8)
2014	813	595 (73.2)	690 (84.9)	751 (92.4)

- The developers noted that sample size likely contributed to the lowest reliability scores.
- The developer also provided the reliability scores for each provider from 2014.

Questions for the Committee:

- *Is the measure score test sample adequate to generalize for widespread implementation?*
- *Do the results from the updated testing 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 /confidence that the performance measure scores are reliable based on the reliability statistic and scope of testing (# of measured entities and representativeness) → 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:

- *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:

- In 2013, face validity was established through a technical work group established for the development of the measures. This measure was found to be important, usable, and feasible by the technical work group overseeing the development of this measure and several others.

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:

- The developer assessed face validity through a technical work group empaneled for the development of the measure. The work group voted on importance, ability to assess quality care, feasibility to implement measure, 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:

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

2b3-2b7. Threats to Validity

2b3. Exclusions:

- N/A

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

- The Ryan White HIV/AIDS Program is a public health, safety net program providing care to a high proportion of racial/ethnic minority, transgender, unstable housing, and low income people living with HIV. Many of people served by the Ryan White HIV/AIDS Program represent sociodemographics factors incorporated in risk adjusting models by many measures stewards. As a result, the Ryan White HIV/AIDS Program does not adjust for risk in its performance measures. Rather, it is a fundamental aspect of the Ryan White HIV/AIDS Program to identify disparities and work to improve quality of care for subpopulations. Additionally, this measure is not used for pay-for-performance, bonuses, or penalties.

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

- The table below demonstrates variability across providers:

	% Patients with viral suppression across providers					Providers achieving ≥80% suppression		
Year	Mean	SD	Median	10th %ile	90th %ile	N	n	%
2010	60.6	23.8	67.8	19.5	82.8	846	145	17.1
2011	64.7	22.1	71.4	31.9	84.9	811	207	24.5
2012	69.9	20.3	75.6	40.2	88.0	816	277	32.7

2013	76.1	17	80.7	57.1	90.2	823	435	51.4
2014	80.3	15.5	84.2	65.0	93.1	813	530	65.2

- In 2014, the bottom 10% of providers had viral suppression rates of 65.0% or lower; the top 90% of providers had viral suppression rates of 93.1% or higher.
- In 2014, of 813 providers, 530 (65.2%) had at least 80% of patients reach viral suppression. Additionally, the overall percentage of patients with viral suppression was 80.3%; however, given the large population that the RWHAP serves, even the poorest performing sites (e.g., bottom 10%) represent a substantial number of patients.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- Missing data analysis not applicable for this measure based on the method used to calculate viral suppression (performance score) because viral suppression status was only calculated for those with count data (<200 copies/mL).
- If no viral load count data were present for a given patient, the patient was considered to have had no viral load test during the measurement year.

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

*Data elements are clearly defined and algorithm is clear.

*"All the data elements are clearly defined
The logic or calculation algorithm is clear
It is likely this measure will be consistently implemented"

*"There is inconsistency in the definition of the denominator statement. in the brief overview, on page 2, the denominator is defined as the number of patients, regardless of age, with a diagnosis of HIV with at lease one medical visit in the measurement year.

In section 2a1, the denominator description provides a few additional conditions, specifically, that patients must be diagnosed with HIV during the first 3 months of the measurement year or prior to the measurement year.

Of note, measure 3210, the corresponding measure does include this condition in the briefed measure information. "

*Reliability was well defined, no concerns at the moment that it can be consistently implemented.

2a2. Reliability Testing
*The sample is large - over 800 providers and 315,00 patients were included. Reliability scores provided are good.

*The measure score test sample is adequate to generalize for widespread implementation
The results from the updated testing demonstrate sufficient reliability so that differences in performance can be identified

*Reliability testing was updated from previous submission. An adequate number of entities were analyzed. Reliability testing demonstrate adequate reliability of the measure using a beta-binomial model to assess the signal to noise ratio.

2b1. Validity Specifications
*The measure is consistent with the evidence

2b2. Validity Testing
*Empirical validity testing was not performed. A technical work group voted on the face validity of the measure. It is not clear to me whether this measure is a better gauge of provider level performance or treatment program level performance.

*"The results demonstrate sufficient validity so that conclusions about quality can be made
My one reservation about this measure as specified is an indicator of quality is that one visit is not enough for a physician to implement a regiment and do follow up testing to see if it worked. Two visits minimum might be better."

*"Reliability testing was performed with an adequate scope. Reliability testing was performed using over 800 ryan white providers and > 315,000 patients. This is adequate to allow for generalizability. Additionally, the measure is not used for pay for performance.

Face validity only was performed. the developer describes a process whereby a technical group and Ryan White recipients were allowed the opportunity to provide feedback. No feedback was garnered that indicated the measure did not have face validity or did not accurately reflect quality. However, there was no systematic process in place to evaluate the ability of the measure to distinguish good and poor quality. "

*There was a wide sampling of population and number of entities and patients. an undetectable viral load is the goal of any HIV therapy, therefore it's the outmost indicator of quality, although there may be reasons for it to be unnatainable in certain populations which are non adherent and do not follow up and get disconnected from care.

2b3-7. Threats to Validity
*Similar to comment in the validity testing section, while there is significant variation in performance, it would be interesting to know if there is significant variation between HIV treatment programs and whether suppression of VL is an indicator or provider quality or program quality.

*"2b.4 Risk adjustment may be necessary for broader application of this measure.
2b.6 The measure identifies meaningful differences about quality"

*There are no exclusions. There is no risk adjustment as the developer provides HIV care to a diverse patient population.

the data presented demonstrate a significant difference in performance across providers during measurement years. In the most recent year reported, 2014, the median % of patients with VL suppression was 80.7%, 10th percentile 57% and 90th at 50%. The poorest performing sites still represent a significant number of patients.

Patients with no viral load or missing values were counted as having no VL testing performed.

*There were no exclusions. Missing data analysis N/A.

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 data elements are generated during the routine provision of care and available in defined fields in electronic health records.
- The data required for operational use of this measure are readily available within patient health records and provided annually to the Ryan White HIV/AIDS Program through the reporting of the Ryan White Service Report (approved by the Office of Management and Budget 0915-0323).
- There are no fees, licensing, or other requirements to use any aspect of 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
*Measure developers have demonstrated feasibility.

*"The required data elements are routinely generated and used during care delivery
The required data elements are available in electronic form, e.g., EHR or other electronic sources"

*all elements are generated during the routine course of care.

*The data is easily accesible in the EMR.

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:

- 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
- Medicaid Adult Core Set
  - Sponsor: Federal government
  - Geographic area: Nationwide
  - Accountable entities: State Medicaid programs
  - Patients: Unknown
- Physician Quality Report System (PQRS) and Value Based Modifier

- Sponsor: Federal government
- Geographic area: Nationwide
- Accountable entities: Physicians and practitioners
- Patients: Unknown

- Merit-Based Incentive Payment System (MIPS)
  - Sponsor: Federal government
  - Geographic area: Nationwide
  - Accountable entities: Physicians, Physician Assistant, Nurse Practitioner, and Clinical Nurse Specialist
  - Patients: Unknown
- National HIV/AIDS Strategy
  - Sponsor: Federal government
  - Geographic area: Nationwide
  - Accountable entities: Federal agencies and service providers
  - Patients: All people living with HIV in the United States

Improvement results:

- The Ryan White HIV/AIDS Program has experienced a 20-point increase in viral suppression from 61.8% in 2010 to 80.3% in 2014. Viral suppression has increased across all demographic groups and subpopulations.

Unexpected findings (positive or negative) during implementation:

- The developer did not provide any unexpected findings during implementation.

Potential harms:

- The developer did not state if any potential harms were identified during implementation of this measure.

Vetting of the measure:

- Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived).
- Ryan White HIV/AIDS Program national partners (national organizations that represent grant recipients, subrecipients, and patients) has provided antidotal feedback regarding the timeliness, feasibility, and usability of the release of the Ryan White HIV/AIDS Program Annual Client-Level Data Report, supplemental reports, slide decks, fact sheets, and infographics. The national partners encourage the continued release of the data in all its formats.
- 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.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*
- *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
*"Data is currently publicly reported and used in 5 accountability programs including Ryan White, State Medicaid programs, PQRS, MIPS and National HIV/AIDS Strategy.

There has been measured improvement in viral suppression rates with implementation of the measure. No unexpected findings have been encountered. "

*In my medical group, this data is given to providers in a monthly basis and helps providers know which patients may have discinnected from care and efforts are put into place to re-engage them. There should be no unintended consequences from measure with caveat that patients are entitled to make poor medical choices and the provider should not be penalized if patient is non adherent with visits or treatment regimen.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0407 HIV/AIDS: HIV RNA Control After Six Months of Potent Antiretroviral Therapy (NCQA)
- 0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis (NCQA)
- 0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis (NCQA)
- 2079 HIV Medical Visit Frequency (HRSA)
- 2080 Gap in HIV Medical Visits (HRSA)
- 2083 Prescription of HIV Antiretroviral Therapy (HRSA)
- 3211 Prescription of HIV Antiretroviral Therapy (HRSA)
- 3210 HIV Viral Suppression (HRSA)
- 3010 HIV Medical Visit Frequency (HRSA)

Harmonization

- Per developer, harmonized with all measures except #0405 and #0409. Plans to harmonize with #0405 and #0409.

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 candidate 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**

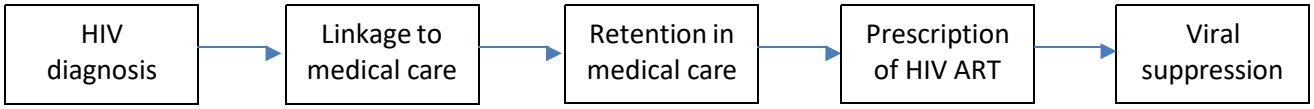
RATIONALE IF NOT ELIGIBLE: The measure is not eligible for Endorsement+ because empirical validity testing of the measure score was not conducted.

Pre-meeting public and member comments

-

Measure Title: Viral Load Supression

1a.12 LOGIC MODEL



Althought the above diagram outlines the sequenctial septs of medical care that people living with HIV go through from initial diagnosis to achieving the goal of viral suppression. For some patients, this is a linear path with sustained viral suppression for many years. For other patients, there many 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 patients 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)

Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. Antiretroviral therapy delays this progression and increases the length of survival.

Antiretroviral therapy 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). Viral suppression is a main goal of HIV treatment and an indicator of treatment success and reduction of potential HIV transmission. It is directly related to:

- Reduction in disease progression, incidence of opportunistic infections, the risk of both defining and non-AIDS-defining complications and the incidence and severity of chronic conditions.
- Reduction in the risk of transmitting HIV to a sexual or drug-using partner who does not have HIV.
- Improvement of immune function, quality of life, increase in time until development of AIDS increase in life expectancy. Being virally suppressed is good for an HIV-positive person's overall health and preventing HIV infection from advancing to AIDS, the last stage of HIV infection.
- Durable viral suppression improves immune function and quality of life, prolongs life, lowers the risk of both AIDS-defining and non-AIDS-defining complications, and prolongs life. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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

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, Accessed November 18, 2016 https://aidsinfo.nih.gov/guidelines/html/1/adult-and-adolescent-arv-guidelines/0

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

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

Günthard HF, Saag MS, Benson CA, del Rio C, Eron JJ, Gallant JE, Hoy JF, Mugavero MJ, Sax PE, Thompson MA, Gandhi RT, Landovitz RJ, Smith DM, Jacobsen DM, Volberding PA. Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults 2016 Recommendations of the International Antiviral Society (IAS)–USA Panel. JAMA. 2016. https://www.iasusa.org/content/antiretroviral-drugs-treatment-and-prevention-hiv-infection-adults-2016-recommendations

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

Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents

Panel's Recommendations for Initiating Antiretroviral Therapy in Treatment-Naive Patients (pE1)
Antiretroviral therapy (ART) is recommended for all HIV-infected individuals, regardless of CD4 T lymphocyte cell count, to reduce the morbidity and mortality associated with HIV infection (**AI**).

Panel's Recommendations for Acute and Recent (Early) HIV Infection (pI1)

- Antiretroviral therapy (ART) is recommended for all individuals with HIV-1 infection **(AI)** including those with early HIV-1 infection.
- Once initiated, the goal of ART is to suppress plasma HIV-1 RNA to undetectable levels **(AIII)**. Testing for plasma HIV-1 RNA levels, CD4 T lymphocyte counts, and toxicity monitoring should be performed as recommended for patients with chronic HIV-1 infection **(AII)**.

Panel's Recommendations Regarding Virologic Failure of the Treatment-Experienced Patient (pH1)

- The goal of treatment for ART-experienced patients with drug resistance who are experiencing virologic failure is to establish virologic suppression (i.e., HIV RNA below the lower limits of detection of currently used assays) **(AI)**.
- Discontinuing or briefly interrupting therapy may lead to a rapid increase in HIV RNA and a decrease in CD4 cell count and increases the risk of clinical progression. Therefore, this strategy is **not** recommended in the setting of virologic failure **(AI)**.

Laboratory Testing, Plasma HIV-1 RNA (Viral Load) and CD4 Count Monitoring (pC5)

- Viral load is the most important indicator of initial and sustained response to ART (AI) and should be measured in all HIV-infected patients at entry into care (AIII), at initiation of therapy (AIII), and on a regular basis thereafter. For those patients who choose to delay therapy, repeat viral load testing while not on ART is optional (CIII).
- Plasma viral load should be measured before initiation of ART and within 2 to 4 weeks but no later than 8 weeks after treatment initiation or modification (AIII). The purpose of the measurements is to confirm an adequate initial virologic response to ART, indicating appropriate regimen selection and patient adherence to therapy. Repeat viral load measurement should be performed at 4- to 8-week intervals until the level falls below the assay's limit of detection (BIII).
- In virologically suppressed patients in whom ART was modified because of drug toxicity or for regimen simplification. Viral load measurement should be performed within 4 to 8 weeks after changing therapy (AIII). The purpose of viral load monitoring at this point is to confirm the effectiveness of the new regimen.
- 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).

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

https://aidsinfo.nih.gov/guidelines/html/3/perinatal-guidelines/0

Panel's Recommendations for HIV-Infected Pregnant Women Who Have Never Received Antiretroviral Drugs (pC27)

- All HIV-infected pregnant women should receive combination antiretroviral therapy (cART) to reduce the risk of perinatal transmission of HIV **(AI)**. The choice of regimen should take into account current adult treatment

guidelines, what is known about the use of specific drugs in pregnancy, and the risk of teratogenicity (see Table 6 and Table 7).

- Consideration should be given to initiating cART as soon as HIV is diagnosed during pregnancy; earlier viral suppression is associated with lower risk of transmission. This decision may be influenced by CD4 T lymphocyte count, HIV RNA levels, and maternal conditions (e.g., nausea and vomiting) **(AIII)**. The benefits of early cART must be weighed against potential fetal effects of drug exposure.

Panel's Recommendations Regarding Lack of Viral Suppression During Antepartum Care (pC48)

- Because maternal antenatal viral load correlates with risk of perinatal transmission of HIV, suppression of HIV RNA to undetectable levels should be achieved as rapidly as possible **(AII)**.
- If an ultrasensitive HIV RNA assay indicates failure of viral suppression (after an adequate period of treatment):
    - Assess adherence and resistance (if HIV RNA level is high enough for resistance testing) **(AII)**.
    - Consult an HIV treatment expert and consider possible antiretroviral regimen modification **(AIII)**.
    - Scheduled cesarean delivery is recommended for HIV-infected pregnant women who have HIV RNA levels >1,000 copies/mL near the time of delivery **(AII)**.

International Advisory Panel on HIV Care Continuum Optimization

- Where possible, jurisdictions should consider longitudinal cohort measurement of HIV service utilization and treatment outcomes to identify the means to maximize viral suppression through ensuring early access to ART and retention in care. (A IV) (p4)

World Health Organization:

- ARV drugs play a key role in HIV prevention. People taking ART who achieve optimal viral suppression are extremely unlikely to pass HIV to sexual partners. ARV drugs taken by people without HIV as PrEP or PEP are highly effective in preventing HIV acquisition. (p64)
- People starting treatment and carergivers should be informed that the first ART regimen offers the best opportunity for effective viral suppression, immune recovery and consequently clinical benefit and that successful ART requires all medications to be taken as prescribed. (p72)
- Access to ART should be the first priority for all age groups, and lack of testing for monitoring treatment response should not be a barrier to initiating ART. If viral load testing capacity is limited, it should be introduced in a phased approach. Examples of phased approaches include: …
    - using viral load initially as a targeted test to confirm treatment failure;
    - prioritizing viral load testing for pregnant and breastfeeding women, especially around the time of delivery, as sustained viral suppression is critical to prevention of transmission to the child, and documented high viral load at delivery is an indication for enhanced infant prophylaxis; (p134)

Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults, 2016 Recommendations of the International Antiviral Society–USA Panel

- HIV RNA level should be monitored every 4 to 6 weeks after treatment is initiated or changed until virus is undetectable (evidence rating AIa). (Box 5)
- After viral suppression is achieved, HIV RNA should be monitored every 3 months until suppressed for 1 year and at least every 6 months thereafter for adherent patients who remain clinically stable (evidence rating AIII). (Box 5)
- When virus has been suppressed for at least 2 years and CD4 cell count is persistently above 500/µL, repeat monitoring of CD4 cell count is not recommended unless virologic failure (evidence rating AIIa) or intercurrent immunosuppressive conditions occur or immunosuppressive treatments are initiated (evidence rating AIII). (Box 5)
- If the HIV RNA level remains above the limit of quantification by 24 weeks after starting new treatment or if rebound above 50 copies/mL occurs at any time, the assay should be repeated within 4 weeks to exclude impending virologic failure (evidence rating AIIa). (Box 5)

- For patients with persistent quantifiable HIV RNA between 50 and 200 copies/mL, reassessment for causes of virologic failure, evaluation again within 4 weeks, and close monitoring are recommended (evidence rating BIII). (Box 5)

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	**I:** One or more randomized trials with clinical outcomes and/or validated laboratory endpoints
B: Moderate recommendation for the statement	**II:** One or more well-designed, non-randomized trials or observational cohort studies with long-term clinical outcomes
C: Optional recommendation for the statement	**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.

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

Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults, 2016 Recommendations of the International Antiviral Society–USA Panel

Table 1. Strength of Recommendation and Quality of Evidence Rating Scale

Rating	Definition
Strength of recommendation	
A	Strong support for the recommendation
B	Moderate support for the recommendation
C	Limited support for the recommendation
Quality of evidence	
Ia	Evidence for $\geq$ 1 randomized clinical trials published in the peer-reviewed literature
Ib	Evidence for $\geq$ 1 randomized clinical trials presented in abstract form at peer-reviewed scientific meetings
IIa	Evidence from nonrandomized clinical trials or cohorts or case-control studies published in the peer-reviewed literature
IIb	Evidence from nonrandomized clinical trials or cohorts or case-control studies published in the peer-reviewed scientific meeting
III	Recommendation based on panel's analysis of the accumulated available evidnce

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

VS_evidence_NQF.docx,VS_submission_form-636179045901415203.docx

1a.1 <u>For Maintenance of Endorsement:</u> 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)*

<u>IF a PRO-PM</u> (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.)

<u>IF a COMPOSITE</u> (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.

Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. Antiretroviral therapy delays this progression and increases the length of survival.

Antiretroviral therapy 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). Viral suppression is a main goal of HIV treatment and an indicator of treatment success and reduction of potential HIV transmission. It is directly related to:

- Reduction in disease progression, incidence of opportunistic infections, the risk of both defining and non-AIDS- defining complications and the incidence and severity of chronic conditions.
- Reduction in the risk of transmitting HIV to a sexual or drug-using partner who does not have HIV.
- Improvement of immune function, quality of life, increase in time until development of AIDS increase in life expectancy.

Being virally suppressed is good for an HIV-positive person's overall health and preventing HIV infection from advancing to AIDS, the last stage of HIV infection.

- Durable viral suppression improves immune function and quality of life, prolongs life, lowers the risk of both AIDS-defining and non-AIDS-defining complications, and prolongs life. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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.*
See attachment "VS 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 attchment "VS 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.)*
https://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:** VS_data_dictionary.docx

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 with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year

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 had a HIV viral load less than 200 copies/mL at the last HIV viral load test during the measurement year

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 measurement year

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 year
2. Patients diagnosed with HIV during the first 3 months of the measurement year or prior to the measurement year
3. Patients who had at least one medical visit during the measurement year

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
There are no patient exclusions.

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.)*
There are no patient exclusions.

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 a HIV during the first 3 months of the measurement year or prior to the measurement year; and 2.) had at least one medical visit during the measurement year. The individuals who met these criteria are the denominator population.
2. Identify the individuals from the denominator population who meet the criterion for inclusion in the numerator: had a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year.
3. Calculate the percentage 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.
Laboratory, 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.
Not applicable.

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
VS_testing-636179043035640093.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

PREVIOUS TESTING FROM INITIAL ENDORSEMENT

NQF #: 2082 **NQF Project:** Infectious Disease Project

2. RELIABILITY & 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. **(evaluation criteria)**
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
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 18 participating treatment sites. However, for this work, we included 9/18 sites. Sites that exclusively used the ultrasensitive and b-DNA HIV viral load assays were included. 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 year included calendar year 2010.

All of the patients in the HIVRN dataset have a diagnosis of HIV. Patients were included, regardless of age, if they had at least one medical visit during the measurement year. For calendar year 2010, 11,134 patients were included in the analysis. The patient characteristics for calendar year 2010 are as follows. The patient characteristics are representative of CDC surveillance data for people living with HIV in 2009 (Table 15a in http://www.cdc.gov/hiv/surveillance/resources/reports/2010report/index.htm).

	2010
Race/Ethnicity:	
African American/Caribbean	46.87%
White, not Hispanic	28.34%
Hispanic	23.06%
Other	1.73%
Gender:	
Male	69.99%
Female	29.26%
Transgender	0.75%
Age:	
<18	2.11%
18-29	11.70%
30-49	56.98%
50+	29.21%
HIV Risk:	
IV Drug Use	12.40%
Men Having Sex with Men	41.59%
Heterosexual Contact	39.44%
Vertical	2.60%
Blood	0.64%
Other/Unknown	3.33%
Insurance:	
Private	18.01%

Medicaid	33.50%
Medicare	15.26%
Dual (Medicare and Medicaid)	1.74%
Uninsured	2.88%
Ryan White	26.70%
Other/Unknown	1.90%

Site Type:

Hospital-based	65.12%
Community-based	34.88%

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)***:**

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 "HIV viral load suppression" 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 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**

Table 1: Clinic-Specific Reliability for Viral Suppression Measure – Year 2010

Between-clinic variance: 0.0066

Clinic	n	percent	Reliability
A	700	77.9	0.96
B	808	50.4	0.96
C	438	78.8	0.95
D	1586	69.9	0.98
E	1739	76.1	0.98
F	4116	72.4	0.99
G	1337	78.5	0.98
Peds	410	67.8	0.93

Median 0.97 (Range 0.93-0.99)

2b. VALIDITY. Validity, Testing, <u>including all Threats to Validity</u>: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* **are consistent with the evidence cited in support of the measure focus (***criterion 1c)* **and identify any differences from the evidence:**

Studies show that lack of HIV viral load suppression leads to poorer health outcomes among people living with HIV. The measure specifications presented are consistent with the elements of HIV viral load suppression as descried in the studies.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

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 18 participating treatment sites. However, for this work, we

included 9/18 sites. Sites that exclusively used the ultrasensitive and b-DNA HIV viral load assays were included. 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 year included calendar years 2010.

All of the patients in the HIVRN dataset have a diagnosis of HIV. Patients were included, regardless of age, if they had at least one medical visit during the measurement year. For calendar year 2010, 11,134 patients were included in the analysis. The patient characteristics for calendar year 2010 are as follows. The patient characteristics are representative of CDC surveillance data for people living with HIV in 2009 (Table 15a in http://www.cdc.gov/hiv/surveillance/resources/reports/2010report/index.htm).

	2010
Race/Ethnicity:	
African American/Caribbean	46.87%
White, not Hispanic	28.34%
Hispanic	23.06%
Other	1.73%
Gender:	
Male	69.99%
Female	29.26%
Transgender	0.75%
Age:	
<18	2.11%
18-29	11.70%
30-49	56.98%
50+	29.21%
HIV Risk:	
IV Drug Use	12.40%
Men Having Sex with Men	41.59%
Heterosexual Contact	39.44%
Vertical	2.60%
Blood	0.64%
Other/Unknown	3.33%
Insurance:	
Private	18.01%
Medicaid	33.50%
Medicare	15.26%
Dual (Medicare and Medicaid)	1.74%
Uninsured	2.88%
Ryan White	26.70%
Other/Unknown	1.90%
Site Type:	
Hospital-based	65.12%
Community-based	34.88%

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:
Face validity was established through a technical work group established for the development of the measures. The technical work group consisted of leading researchers and physicians in HIV retention, care, and treatment as well as governmental and non-governmental 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. Often, the principle investigator of the study presented to the work group. 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, feasibilityn use in quality improvement activities. The votes were tallied and draft components of the measures 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.

Additional face validity was gained through a structured process of webinar presentations to a national audience of Ryan White Program providers. The Ryan White providers were presented detailed information about each of the measures (e.g. work group process, numerator, demoninator, exclusions, etc.) via a webinar. After receiving the detailed information about the measures, Ryan White providers were asked to implement the measures within their quality management program and provide feedback on the feasibility and usability of the measures. Feedback was gathered during an additional webinar and written responses.

2b2.3 **Testing Results** *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)***:**
This measure was found to be important, usable, and feasible by the technical work group overseeing the development of this measure and several others. The technical work group considered 7 measures. In total, 4 of the 7 measures were voted as the most import, feasible, and useable. The Ryan White providers have also deemed the measures important, usable, and feasible. Over 180 Ryan White providers from across the country have voluntary reported performance data for this measure at least once with 148 of those providers reporting performance data for 4 straight measurement periods.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 **Data/Sample for analysis of exclusions** *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**
Not applicable.

2b3.2 **Analytic Method** (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)***:**
Not applicable.

2b3.3 **Results** *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
Not applicable.

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 **Data/Sample** *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**
Not applicable.

2b4.2 **Analytic Method (***Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)***:**
Not applicable.

2b4.3 **Testing Results** *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)***:**
Not applicable.

2b4.4 **If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:**
Not applicable.

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

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 18 participating treatment sites. However, for this work, we included 9/18 sites. Sites that exclusively used the ultrasensitive and b-DNA HIV viral load assays were included. 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 year included calendar years 2010.

All of the patients in the HIVRN dataset have a diagnosis of HIV. Patients were included, regardless of age, if they had at least one medical visit during the measurement year. For calendar year 2010, 11,134 patients were included in the analysis. The patient characteristics for calendar year 2010 are as follows. The patient characteristics are representative of CDC surveillance data for people living with HIV in 2009 (Table 15a in http://www.cdc.gov/hiv/surveillance/resources/reports/2010report/index.htm).

	2010
Race/Ethnicity:	
African American/Caribbean	46.87%
White, not Hispanic	28.34%
Hispanic	23.06%
Other	1.73%
Gender:	
Male	69.99%
Female	29.26%
Transgender	0.75%
Age:	
<18	2.11%
18-29	11.70%
30-49	56.98%
50+	29.21%
HIV Risk:	
IV Drug Use	12.40%
Men having sex with Men	41.59%
Heterosexual Contact	39.44%
Vertical	2.60%
Blood	0.64%
Other/Unknown	3.33%
Insurance:	
Private	18.01%
Medicaid	33.50%
Medicare	15.26%
Dual (Medicare and Medicaid)	1.74%
Uninsured	2.88%
Ryan White	26.70%
Other/Unknown	1.90%
Site Type:	
Hospital-based	65.12%
Community-based	34.88%

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)***:**

We reported the mean, minimum, maximum, and percentile.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
Among the 9 sites (2 pediatric sites were combined due to small patient populations), the following data are reported for measurement year 2010.
Minimum 50.37%
Maximum 78.77%
Mean 72.20%
25th percentile 69.39%
50th percentile 74.29%
75th percentile 78.01%

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure was not tested with multiple data sources.

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:
This measure was not tested with multiple data sources.

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:
This measure was not tested with multiple data sources.

2c. Disparities in Care: H☐ M☐ L☐ I☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: The following are the results stratified by patient characteristics and site.
2010
Race/Ethnicity:
African American/Caribbean 67.16%
White, not Hispanic 78.84%
Hispanic 74.16%
Other 71.73%

Gender:
Male 74.22%
Female 67.56%
Transgender 65.06%

Age:
<18 78.30%
18-29 56.10%
30-49 71.19%
50+ 80.20%

HIV Risk:
IV Drug Use 69.44%
Men Having Sex with Men 76.31%
Heterosexual Contact 69.39%
Vertical 75.43%
Blood 69.01%
Other/Unknown 62.53%

Insurance:
Private 76.21%
Medicaid 68.69%
Medicare 73.16%
Dual (Medicare and Medicaid) 81.96%
Uninsured 55.45%
Ryan White 74.87%
Other/Unknown 67.45%

Site Type:
Hospital-based 72.54%
Community-based 71.58%

Site:
A 77.86%
B 50.37%
C 78.77%
D 69.92%
E 76.14%
F 72.45%
G 78.46%
Pediatric Sites (combined) 67.80%

2c.2 **If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:**
Not applicable

2.1-2.3 **Supplemental Testing Methodology Information:**

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met? (*Reliability and Validity must be rated moderate or high*) Yes ☐ No ☐
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

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

Measure Number (*if previously endorsed*)**:** 2082
Measure Title: HIV viral suppression
Date of Submission: Click here to enter a date
Type of Measure:

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

☐ 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: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
On an annual basis, Ryan White HIV/AIDS Program (RWHAP) grant recipient and subrecipients submit the Ryan White HIV/AIDS Services Report (RSR). The RSR dataset is the Health Resources and Services Administration HIV/AIDS Bureau's primary source of annual, client-level data collected from its nearly 2,000 funded grant recipients and subrecipients. Since 2010, client-level RSR data have been used to assess the numbers and types of clients receiving services and their HIV outcomes. Project Officers at the HIV/AIDS Bureau share the data with grant recipients and subrecipients to monitor and support their progress at improving care and treatment for people living with HIV. It is through the hard work of these providers and the RWHAP community that clients are helped every day.

RSR includes all clients served by the RWHAP during calendar years 2010 through 2014. RSR data do not include information about AIDS Drug Assistance Programs (ADAP); all ADAP-related information is collected through another data system. Although data presented in this report are "nonADAP," this does not imply the clients did not receive ADAP services. ADAP data will be published separately, at later time.

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:	**Measure Tested at Level of:**

(*must be consistent with levels entered in item S.26*)	
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

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

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. Over 800 (varies by year) Ryan White HIV/AIDS Program outpatient ambulatory medical care providers representing various types, locations, and sizes were included in the testing.

Descriptive characteristics of RWHAP providers

	2010		2011		2012		2013		2014	
	N	%	N	%	N	%	N	%	N	%
Overall	846		811		816		823	--	813	--
Provider type										
Hospital or university-based clinic	355	17.5	358	18.6	349	19.1	351	19.6	338	19.4
Community based organization	1,114	54.9	1,053	54.8	993	54.3	958	53.6	921	53.0
Health department	284	14.0	274	14.3	243	13.3	233	13.0	243	14.0
Other	275	13.6	237	12.3	243	13.3	247	13.8	237	13.6
HHS Region										
Region 1	149	8.0	153	8.6	142	8.4	139	8.4	135	8.3
Region 2	368	19.7	339	19.0	323	19.1	303	18.3	293	18.1
Region 3	180	9.6	177	9.9	174	10.3	174	10.5	160	9.9
Region 4	337	18.0	335	18.8	312	18.5	301	18.1	313	19.3
Region 5	197	10.5	189	10.6	177	10.5	188	11.3	180	11.1
Region 6	150	8.0	142	8.0	133	7.9	131	7.9	132	8.2
Region 7	65	3.5	60	3.4	57	3.4	56	3.4	54	3.3
Region 8	48	2.6	43	2.4	34	2.0	35	2.1	46	2.8
Region 9	300	16.0	281	15.7	277	16.4	276	16.6	253	15.6
Region 10	78	4.2	68	3.8	60	3.6	56	3.4	52	3.2

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.

Distribution of patients per provider by year, 2010-2014

Year	N patients, mean	N patients, median	Min patients	Max patients
2010	384	177	1	13,159
2011	404	182	1	13,380
2012	411	179	1	13,849
2013	398	181	1	14,755
2014	388	177	1	13,850

Descriptive characteristics of RWHAP patients by year, 2010-2014

	2010		2011		2012		2013		2014	
	No.	%	No.	%	No.	%	No.	%	No.	%
OVERALL	324,455	—	327,744	—	335,408	—	327,618	—	316,087	—
AGE GROUP										
<13	3,709	1.2	3,647	1.1	3,150	1.0	2,667	0.9	2,720	0.9
13–14	627	0.2	605	0.2	469	0.1	360	0.1	343	0.1
15–19	3,698	1.2	3,541	1.1	3,066	0.9	2,609	0.8	2,506	0.8
20–24	14,040	4.5	14,831	4.6	15,741	4.8	15,538	5.0	14,578	4.8
25–29	22,120	7.0	23,278	7.3	24,904	7.7	25,586	8.2	26,043	8.5
30–34	28,644	9.1	29,330	9.2	30,084	9.3	29,495	9.4	28,484	9.3
35–39	35,161	11.2	33,597	10.5	33,005	10.2	31,560	10.1	30,691	10.0
40–44	50,769	16.1	47,941	15.0	45,343	14.0	40,728	13.0	37,000	12.1
45–49	60,344	19.2	59,453	18.6	58,145	17.9	52,863	16.8	47,932	15.6
50–54	46,433	14.7	48,647	15.2	50,876	15.7	50,491	16.1	50,492	16.4
55–59	28,015	8.9	30,646	9.6	33,215	10.2	33,493	10.7	34,667	11.3
60–64	13,441	4.3	15,237	4.8	16,991	5.2	17,780	5.7	19,399	6.3
≥65	8,187	2.6	8,946	2.8	10,147	3.1	10,780	3.4	12,231	4.0
RACE/ETHNICITY										
American Indian/ Alaska Native	1,473	0.5	1,366	0.4	1,371	0.4	1,414	0.5	1,272	0.4
Asian	3,382	1.1	3,598	1.2	3,980	1.2	3,835	1.2	3,791	1.2
Black/ African American	146,460	47.3	149,834	47.8	150,974	47.2	146,056	47.0	142,746	46.9
Hispanic/Latino[a]	71,002	22.9	71,240	22.7	75,201	23.5	74,967	24.1	74,714	24.5
Native Hawaiian/ Pacific Islander	627	0.2	710	0.2	575	0.2	510	0.2	442	0.2
White	83,854	27.1	83,061	26.5	83,820	26.2	78,953	25.4	75,931	24.9
Multiple races	3,177	1.0	3,716	1.2	4,238	1.3	4,899	1.6	5,651	1.9

GENDER										
Male	219,625	69.7	223,379	69.9	230,075	70.8	221,930	70.7	216,965	70.7
Female	93,266	29.6	93,687	29.3	92,186	28.4	89,212	28.4	87,071	28.4
Transgender	2,313	0.7	2,585	0.8	2,848	0.9	2,779	0.9	2,974	1.0
TRANSMISSION RISK										
Male client										
Male-to-male sexual contact	117,267	59.9	120,622	60.2	128,744	61.8	127,571	62.2	127,624	62.7
Injection drug use	17,479	8.9	16,787	8.4	15,586	7.5	15,509	7.6	13,753	6.8
Male-to-male sexual contact and injection drug use	6,971	3.6	6,837	3.4	6,974	3.3	6,136	3.0	6,396	3.1
Heterosexual contact	48,903	25.0	50,814	25.4	52,266	25.1	51,174	24.9	51,155	25.1
Perinatal infection	3,830	2.0	3,919	2.0	3,604	1.7	3,419	1.7	3,456	1.7
Other	1,248	0.6	1,231	0.6	1,309	0.6	1,402	0.7	1,189	0.6
Female client										
Injection drug use	9,264	11.2	9,022	10.7	8,182	9.8	8,310	10.0	7,396	9.1
Heterosexual contact	68,009	82.4	69,767	82.8	70,362	84.1	69,356	83.9	69,090	84.8
Perinatal infection	4,338	5.3	4,587	5.4	4,182	5.0	4,003	4.8	4,093	5.0
Other	900	1.1	877	1.0	936	1.1	1,044	1.3	940	1.2
Transgender client										
Sexual contact	1,874	90.7	2,058	91.2	2,281	91.8	2,314	92.9	2,499	93.2
Injection drug use	38	1.8	32	1.4	35	1.4	32	1.3	31	1.2
Sexual contact and injection drug use	144	7.0	156	6.9	158	6.4	130	5.2	135	5.0
Perinatal infection	5	0.2	5	0.2	2	0.1	4	0.2	9	0.3
Other	6	0.3	5	0.2	8	0.3	10	0.4	8	0.3
HEALTH CARE COVERAGE										
Private only	35,392	12.4	37,532	12.3	39,972	12.7	37,204	12.1	—	—
Medicare only	23,245	8.1	24,279	8.0	23,538	7.5	22,840	7.5	—	—
Medicaid only	73,292	25.6	75,690	24.8	71,990	22.8	69,211	22.6	—	—
Other public	22,398	7.8	20,977	6.9	28,039	8.9	27,347	8.9	—	—
Other private	11,512	4.0	9,884	3.2	6,049	1.9	3,682	1.2	—	—
No coverage	86,220	30.1	100,001	32.8	103,150	32.7	101,524	33.1	—	—
Multiple coverages	34,276	12.0	36,330	11.9	42,969	13.6	44,578	14.6	—	—
Private employer	—	—	—	—	—	—	—	—	18,805	6.3
Private individual	—	—	—	—	—	—	—	—	16,154	5.4
Medicare	—	—	—	—	—	—	—	—	26,145	8.7
Medicaid	—	—	—	—	—	—	—	—	94,993	31.6
Medicare and Medicaid	—	—	—	—	—	—	—	—	19,207	6.4
Veterans Administration	—	—	—	—	—	—	—	—	454	0.2
Indian Health Service	—	—	—	—	—	—	—	—	71	0.0
Other plan	—	—	—	—	—	—	—	—	11,899	4.0
No coverage	—	—	—	—	—	—	—	—	90,828	30.2
Multiple coverages	—	—	—	—	—	—	—	—	22,428	7.5

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.

Ryan White HIV/AIDS Program Services Report (RSR) 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 (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

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.

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*)

Provider-level reliability results for the "viral load suppression" measure in 2014 are detailed below. Results for years 2010-2013 are available upon request, but were not included due to space constraints.

Provider-level reliability testing (signal to noise) results, 2014.

Site/provider ID	% Suppressed	Variance Within	Reliability

55	67.6%	0.001	0.960
63	41.7%	0.010	0.556
82	89.2%	0.001	0.925
88	84.2%	0.000	0.990
96	68.7%	0.000	0.973
101	85.0%	0.001	0.933
105	82.8%	0.001	0.928
112	92.3%	0.000	0.972
113	95.6%	0.000	0.987
117	88.6%	0.000	0.993
118	94.3%	0.000	0.987
120	84.3%	0.001	0.936
123	86.5%	0.000	0.980
124	73.6%	0.001	0.932
127	69.4%	0.001	0.945
128	84.9%	0.000	0.980
133	74.8%	0.000	0.994
135	93.6%	0.000	0.967
138	70.5%	0.005	0.728
140	100.0%	0.000	1.000
141	67.6%	0.001	0.926
143	91.1%	0.001	0.951
144	85.5%	0.000	0.972
147	71.9%	0.001	0.897
148	65.3%	0.005	0.732
149	89.6%	0.001	0.940
154	93.0%	0.000	0.991
155	87.0%	0.005	0.720
156	52.2%	0.005	0.700
158	91.0%	0.001	0.954
159	89.5%	0.000	0.967
160	75.6%	0.002	0.861
164	77.5%	0.002	0.837
168	90.1%	0.000	0.991
169	87.5%	0.001	0.943
170	90.7%	0.000	0.975
171	82.5%	0.001	0.961
172	96.9%	0.000	0.992
173	81.2%	0.001	0.927

174	92.8%	0.000	0.985
175	86.0%	0.001	0.943
176	87.6%	0.000	0.996
177	84.1%	0.002	0.857
178	86.4%	0.000	0.994
179	90.1%	0.001	0.920
181	85.7%	0.000	0.982
182	50.4%	0.000	0.979
183	88.5%	0.000	0.983
184	93.5%	0.000	0.987
186	95.1%	0.000	0.965
187	90.6%	0.000	0.984
188	76.2%	0.000	0.971
191	90.7%	0.000	0.982
192	76.8%	0.000	0.975
194	82.0%	0.001	0.951
196	87.8%	0.000	0.993
197	83.8%	0.000	0.991
199	77.4%	0.000	0.974
201	85.3%	0.000	0.997
203	85.9%	0.000	0.988
205	66.7%	0.019	0.406
207	89.7%	0.000	0.978
209	76.0%	0.001	0.930
210	76.7%	0.001	0.895
211	81.0%	0.000	0.974
212	82.1%	0.001	0.953
213	76.9%	0.000	0.989
214	90.2%	0.000	0.990
215	81.8%	0.000	0.967
216	85.3%	0.000	0.980
217	86.2%	0.000	0.963
220	90.3%	0.000	0.977
221	83.9%	0.001	0.946
222	86.3%	0.000	0.974
223	78.7%	0.000	0.992
224	90.1%	0.000	0.987
225	83.6%	0.000	0.994
227	81.0%	0.000	0.988

228	87.2%	0.001	0.955
230	85.2%	0.000	0.996
231	70.7%	0.000	0.976
232	71.7%	0.001	0.935
233	77.3%	0.000	0.987
235	93.1%	0.000	0.966
236	90.1%	0.000	0.980
238	85.9%	0.000	0.992
239	74.0%	0.000	0.979
240	74.3%	0.000	0.966
241	74.4%	0.000	0.995
242	87.0%	0.001	0.952
244	76.1%	0.000	0.968
245	89.1%	0.000	0.989
246	84.0%	0.000	0.992
248	85.3%	0.000	0.995
252	87.0%	0.000	0.972
253	92.1%	0.000	0.993
255	60.2%	0.002	0.838
256	74.6%	0.001	0.954
257	93.7%	0.000	0.991
259	82.0%	0.001	0.905
263	76.2%	0.000	0.977
265	85.0%	0.000	0.996
266	89.7%	0.000	0.993
267	89.5%	0.000	0.993
268	69.2%	0.001	0.950
269	87.8%	0.000	0.982
271	93.1%	0.001	0.934
273	82.4%	0.000	0.992
275	87.4%	0.000	0.992
276	91.8%	0.000	0.978
277	84.7%	0.000	0.988
278	80.7%	0.000	0.983
279	88.5%	0.000	0.993
280	87.5%	0.003	0.787
283	89.8%	0.000	0.994
284	87.7%	0.000	0.971
285	72.2%	0.001	0.937

286	84.4%	0.000	0.975
288	83.3%	0.001	0.929
289	78.8%	0.001	0.959
290	85.7%	0.002	0.888
291	83.8%	0.000	0.983
292	86.8%	0.001	0.961
294	74.2%	0.000	0.989
295	77.5%	0.001	0.958
298	93.5%	0.000	0.997
299	92.0%	0.000	0.992
302	90.2%	0.000	0.991
303	87.2%	0.000	0.971
304	82.3%	0.000	0.985
305	86.2%	0.000	0.991
307	86.2%	0.000	0.990
308	92.6%	0.001	0.952
310	81.8%	0.000	0.996
311	60.6%	0.001	0.917
312	90.3%	0.001	0.931
313	80.8%	0.001	0.925
314	88.3%	0.000	0.993
315	87.7%	0.000	0.992
316	88.8%	0.000	0.968
317	88.5%	0.000	0.963
318	95.3%	0.000	0.992
319	90.0%	0.001	0.948
320	95.0%	0.000	0.977
321	80.4%	0.001	0.917
322	80.4%	0.001	0.929
323	89.9%	0.000	0.980
324	94.7%	0.000	0.983
325	76.5%	0.000	0.992
326	93.6%	0.000	0.994
328	79.1%	0.000	0.977
329	86.2%	0.000	0.998
332	88.9%	0.000	0.983
333	80.6%	0.002	0.888
334	89.4%	0.000	0.977
335	87.5%	0.000	0.984

336	90.1%	0.000	0.984
340	81.2%	0.001	0.961
342	86.6%	0.000	0.976
343	83.7%	0.001	0.955
344	76.5%	0.000	0.978
345	80.0%	0.001	0.905
347	86.3%	0.000	0.991
348	88.7%	0.000	0.976
349	80.2%	0.001	0.962
351	91.7%	0.000	0.990
353	86.5%	0.000	0.986
357	82.5%	0.000	0.990
358	85.2%	0.001	0.935
360	84.6%	0.000	0.987
361	83.4%	0.000	0.982
362	85.5%	0.002	0.876
363	81.4%	0.001	0.908
365	95.2%	0.001	0.921
366	89.3%	0.001	0.959
368	86.0%	0.000	0.973
369	91.5%	0.000	0.974
370	82.0%	0.002	0.884
371	76.9%	0.002	0.881
372	82.9%	0.000	0.990
375	85.1%	0.000	0.994
378	90.8%	0.001	0.908
379	88.1%	0.000	0.993
380	60.5%	0.000	0.963
382	89.5%	0.000	0.989
384	82.7%	0.001	0.961
385	87.6%	0.001	0.960
386	91.0%	0.000	0.965
388	89.0%	0.000	0.994
389	85.4%	0.001	0.954
390	90.3%	0.000	0.981
391	79.5%	0.001	0.945
393	91.4%	0.000	0.992
394	96.9%	0.000	0.964
395	87.0%	0.000	0.988

400	82.6%	0.001	0.954
404	90.9%	0.001	0.959
407	35.0%	0.001	0.954
408	86.4%	0.000	0.985
409	85.7%	0.001	0.929
410	63.8%	0.003	0.814
412	79.1%	0.000	0.973
414	85.4%	0.001	0.930
417	76.8%	0.002	0.888
421	92.1%	0.000	0.991
422	86.1%	0.000	0.977
423	83.9%	0.000	0.973
425	74.5%	0.001	0.940
427	89.8%	0.000	0.996
438	76.6%	0.000	0.975
441	68.4%	0.003	0.817
457	92.6%	0.000	0.995
463	70.6%	0.001	0.954
469	73.4%	0.001	0.961
473	0.0%	0.000	1.000
480	84.8%	0.000	0.973
481	76.4%	0.001	0.953
483	69.3%	0.000	0.981
489	69.3%	0.000	0.978
491	3.4%	0.000	0.972
498	92.0%	0.000	0.990
504	86.5%	0.000	0.994
506	89.1%	0.000	0.976
509	85.1%	0.000	0.987
510	80.6%	0.001	0.910
517	72.2%	0.001	0.952
534	93.3%	0.000	0.990
553	17.4%	0.006	0.670
593	90.2%	0.001	0.946
598	79.1%	0.001	0.949
612	78.9%	0.002	0.872
664	73.5%	0.001	0.954
704	92.9%	0.000	0.976
710	68.6%	0.004	0.750

726	49.2%	0.001	0.938
738	72.2%	0.011	0.532
744	78.2%	0.000	0.968
753	79.8%	0.001	0.960
757	87.2%	0.001	0.938
762	74.4%	0.000	0.977
765	81.3%	0.001	0.954
775	86.8%	0.001	0.921
783	60.0%	0.005	0.704
787	88.9%	0.001	0.951
791	93.4%	0.000	0.985
793	85.7%	0.006	0.685
794	68.2%	0.003	0.794
798	84.9%	0.000	0.987
799	90.7%	0.002	0.891
800	82.4%	0.001	0.943
801	93.3%	0.002	0.859
803	79.7%	0.000	0.966
807	83.8%	0.001	0.956
818	85.9%	0.000	0.993
820	84.0%	0.001	0.899
821	91.5%	0.001	0.930
824	81.8%	0.003	0.789
841	82.8%	0.002	0.892
852	79.9%	0.001	0.916
861	82.2%	0.001	0.952
867	51.2%	0.000	0.979
871	89.0%	0.000	0.994
873	79.9%	0.001	0.956
894	81.9%	0.000	0.994
905	74.1%	0.000	0.983
907	80.0%	0.000	0.996
913	79.7%	0.000	0.990
920	66.7%	0.015	0.461
926	81.1%	0.004	0.753
927	88.5%	0.001	0.934
929	90.7%	0.001	0.955
933	79.4%	0.001	0.940
945	88.4%	0.001	0.937

980	87.3%	0.001	0.942
986	85.1%	0.001	0.934
992	86.9%	0.000	0.993
996	89.0%	0.000	0.991
1009	55.6%	0.002	0.887
1017	83.0%	0.002	0.894
1022	86.0%	0.001	0.962
1023	87.7%	0.001	0.958
1026	79.0%	0.002	0.889
1029	62.9%	0.001	0.896
1031	89.0%	0.000	0.986
1036	87.2%	0.000	0.977
1037	87.5%	0.001	0.923
1038	78.2%	0.001	0.959
1049	71.4%	0.000	0.964
1050	77.8%	0.000	0.984
1052	85.2%	0.005	0.730
1055	83.3%	0.000	0.969
1056	87.7%	0.000	0.989
1066	90.8%	0.000	0.989
1067	85.2%	0.000	0.985
1068	67.5%	0.000	0.988
1093	73.9%	0.003	0.819
1094	86.5%	0.000	0.988
1100	83.5%	0.000	0.981
1109	89.1%	0.000	0.993
1110	81.7%	0.001	0.910
1112	84.9%	0.000	0.994
1120	88.9%	0.000	0.990
1121	87.1%	0.000	0.995
1122	82.5%	0.000	0.973
1131	83.8%	0.000	0.990
1132	62.9%	0.001	0.910
1146	73.3%	0.013	0.493
1155	98.6%	0.000	0.998
1160	66.1%	0.000	0.987
1162	75.9%	0.003	0.789
1163	77.2%	0.001	0.956
1167	84.7%	0.000	0.996

1214	96.6%	0.000	0.983
1216	6.5%	0.001	0.958
1229	88.0%	0.000	0.971
1230	90.4%	0.001	0.955
1263	100.0%	0.000	1.000
1276	83.9%	0.000	0.968
1278	83.5%	0.001	0.904
1284	86.7%	0.001	0.915
1287	87.3%	0.000	0.990
1289	81.6%	0.004	0.762
1300	37.1%	0.004	0.771
1302	1.8%	0.000	0.996
1309	61.1%	0.003	0.793
1310	94.6%	0.001	0.933
1314	36.2%	0.000	0.967
1318	79.5%	0.002	0.850
1319	82.1%	0.002	0.891
1333	84.7%	0.000	0.965
1349	61.4%	0.004	0.753
1358	58.3%	0.001	0.901
1359	68.8%	0.000	0.992
1364	78.1%	0.002	0.844
1378	84.4%	0.001	0.931
1380	79.5%	0.002	0.858
1382	53.3%	0.002	0.875
1401	83.4%	0.000	0.982
1430	62.3%	0.004	0.741
1444	83.8%	0.004	0.775
1445	82.1%	0.000	0.993
1448	85.7%	0.006	0.685
1451	80.9%	0.002	0.885
1456	84.1%	0.000	0.994
1461	93.7%	0.000	0.965
1464	52.4%	0.001	0.933
1479	83.3%	0.012	0.522
1490	35.0%	0.006	0.690
1511	85.5%	0.000	0.985
1512	61.9%	0.001	0.907
1514	79.1%	0.004	0.767

1527	65.0%	0.011	0.527
1552	90.8%	0.000	0.996
1567	85.6%	0.001	0.902
1570	74.5%	0.001	0.911
1572	70.6%	0.004	0.757
1574	80.2%	0.002	0.884
1582	90.1%	0.001	0.958
1583	75.0%	0.047	0.213
1587	91.7%	0.006	0.665
1594	83.1%	0.002	0.841
1597	81.5%	0.006	0.694
1607	85.8%	0.001	0.960
1610	87.5%	0.001	0.962
1628	96.9%	0.001	0.930
1634	79.8%	0.002	0.891
1635	83.6%	0.001	0.911
1637	79.9%	0.001	0.961
1650	93.7%	0.000	0.964
1654	73.1%	0.004	0.770
1656	90.2%	0.001	0.950
1668	80.5%	0.000	0.988
1672	96.3%	0.001	0.906
1684	79.7%	0.000	0.993
1719	80.2%	0.000	0.990
1762	79.3%	0.000	0.986
1784	85.3%	0.001	0.940
1786	87.9%	0.002	0.887
1792	78.1%	0.000	0.984
1806	55.6%	0.027	0.316
1809	84.8%	0.000	0.986
1812	81.0%	0.000	0.984
1831	93.3%	0.001	0.939
1834	71.4%	0.005	0.723
1847	79.6%	0.002	0.884
1849	79.9%	0.001	0.919
1879	80.1%	0.001	0.938
1900	0.0%	0.000	1.000
1904	78.6%	0.006	0.678
1912	88.0%	0.000	0.987

1930	81.8%	0.001	0.903
1955	86.2%	0.001	0.925
1967	69.7%	0.006	0.664
1968	90.6%	0.003	0.827
1970	100.0%	0.000	1.000
1972	0.0%	0.000	1.000
1977	92.6%	0.003	0.833
1980	79.6%	0.003	0.793
1989	85.2%	0.001	0.916
2003	50.0%	0.063	0.168
2008	100.0%	0.000	1.000
2010	100.0%	0.000	1.000
2011	81.0%	0.007	0.633
2017	87.9%	0.003	0.797
2020	90.0%	0.009	0.585
2025	86.2%	0.000	0.966
2028	85.1%	0.001	0.951
2029	87.5%	0.001	0.911
2034	92.5%	0.001	0.906
2041	75.0%	0.023	0.351
2049	75.6%	0.005	0.738
2058	88.0%	0.000	0.972
2072	87.9%	0.003	0.797
2073	91.1%	0.000	0.991
2076	100.0%	0.000	1.000
2078	84.4%	0.004	0.754
2080	100.0%	0.000	1.000
2081	100.0%	0.000	1.000
2116	80.7%	0.001	0.899
2117	69.6%	0.009	0.579
2118	82.8%	0.002	0.892
2126	87.4%	0.001	0.940
2127	84.4%	0.001	0.897
2129	82.1%	0.001	0.924
2133	86.2%	0.000	0.974
2134	74.2%	0.006	0.672
2137	88.5%	0.004	0.763
2139	6.1%	0.000	0.973
2141	81.0%	0.000	0.966

2143	89.2%	0.000	0.992
2148	88.7%	0.001	0.940
2150	87.4%	0.000	0.994
2153	84.2%	0.001	0.938
2163	87.8%	0.001	0.960
2170	78.6%	0.001	0.960
2174	79.4%	0.000	0.998
2175	88.0%	0.001	0.958
2178	78.1%	0.000	0.968
2180	84.7%	0.000	0.990
2183	93.1%	0.000	0.972
2187	95.8%	0.002	0.884
2188	94.7%	0.003	0.828
2189	90.5%	0.004	0.755
2191	91.9%	0.001	0.914
2198	87.8%	0.002	0.852
2200	84.8%	0.004	0.765
2203	82.6%	0.001	0.906
2205	88.7%	0.002	0.870
2207	78.9%	0.000	0.994
2224	87.3%	0.000	0.988
2228	36.8%	0.012	0.508
2230	90.3%	0.001	0.942
2232	60.0%	0.048	0.209
2246	74.2%	0.001	0.939
2252	18.8%	0.002	0.842
2263	92.0%	0.003	0.811
2264	50.0%	0.125	0.092
2296	88.4%	0.000	0.974
2299	79.6%	0.000	0.977
2320	89.2%	0.000	0.992
2366	90.0%	0.001	0.908
2368	97.1%	0.000	0.968
2374	100.0%	0.000	1.000
2378	89.5%	0.005	0.719
2379	91.3%	0.003	0.786
2381	94.8%	0.000	0.976
2388	66.7%	0.037	0.255
2389	0.0%	0.000	1.000

2415	96.1%	0.000	0.978
2420	87.9%	0.001	0.937
2436	92.3%	0.000	0.970
2438	93.2%	0.001	0.922
2444	50.0%	0.125	0.092
2457	94.2%	0.000	0.981
2474	68.1%	0.001	0.939
2495	92.5%	0.001	0.925
2514	39.7%	0.000	0.999
2525	91.0%	0.000	0.992
2572	76.5%	0.000	0.978
2654	94.7%	0.001	0.906
2694	100.0%	0.000	1.000
2699	94.7%	0.003	0.828
2700	87.8%	0.000	0.981
2702	86.9%	0.000	0.979
2703	89.9%	0.000	0.986
2704	87.1%	0.000	0.974
2707	100.0%	0.000	1.000
2709	85.6%	0.000	0.991
2714	76.4%	0.001	0.947
2717	85.6%	0.001	0.928
2718	88.1%	0.000	0.996
2720	77.0%	0.001	0.900
2721	70.3%	0.000	0.985
2723	83.6%	0.000	0.970
2728	93.9%	0.000	0.983
2732	88.2%	0.000	0.981
2737	72.1%	0.005	0.730
2745	93.3%	0.001	0.961
2746	83.3%	0.001	0.920
2757	85.8%	0.001	0.959
2764	71.1%	0.001	0.920
2766	76.3%	0.001	0.916
2767	73.8%	0.005	0.733
2768	82.8%	0.000	0.993
2769	60.3%	0.004	0.769
2771	77.6%	0.001	0.961
2775	88.0%	0.000	0.989

2779	85.8%	0.001	0.959
2782	72.0%	0.001	0.922
2788	78.3%	0.001	0.939
2790	71.1%	0.000	0.968
2794	90.4%	0.002	0.883
2795	80.6%	0.001	0.916
2849	89.0%	0.000	0.980
2855	82.4%	0.001	0.920
2856	76.4%	0.003	0.794
2857	100.0%	0.000	1.000
2865	87.4%	0.000	0.985
2872	97.6%	0.001	0.956
2873	57.1%	0.035	0.266
2892	75.1%	0.001	0.949
2893	88.7%	0.000	0.971
2928	93.1%	0.000	0.990
2929	85.9%	0.001	0.899
2933	90.4%	0.000	0.963
2945	66.7%	0.001	0.923
2947	82.9%	0.000	0.986
2948	81.7%	0.000	0.976
2949	67.9%	0.001	0.904
2951	82.9%	0.001	0.948
2958	91.7%	0.002	0.856
2959	57.7%	0.005	0.730
2966	75.8%	0.002	0.892
2969	72.9%	0.003	0.818
2970	94.0%	0.001	0.957
2988	57.0%	0.003	0.828
2989	77.8%	0.019	0.397
2997	81.9%	0.000	0.987
3010	77.2%	0.001	0.962
3012	83.1%	0.001	0.943
3020	100.0%	0.000	1.000
3048	73.7%	0.010	0.554
3052	75.6%	0.001	0.902
3077	70.9%	0.003	0.829
3079	91.8%	0.001	0.957
3111	90.1%	0.001	0.910

3131	79.7%	0.001	0.943
3133	74.2%	0.002	0.855
3164	88.0%	0.002	0.857
3177	89.7%	0.001	0.958
3187	69.9%	0.003	0.814
3210	75.2%	0.001	0.955
3255	54.7%	0.003	0.793
3261	86.1%	0.001	0.941
3262	70.5%	0.005	0.728
3264	75.9%	0.006	0.667
3265	78.0%	0.001	0.929
3310	72.6%	0.001	0.926
3359	65.6%	0.002	0.875
3389	66.7%	0.007	0.653
3397	39.1%	0.001	0.926
3401	35.7%	0.002	0.861
3428	92.4%	0.000	0.988
3433	79.0%	0.001	0.940
3440	88.0%	0.000	0.970
3444	85.4%	0.001	0.956
3445	85.7%	0.002	0.853
3449	70.5%	0.005	0.728
3456	61.2%	0.001	0.949
3469	87.1%	0.000	0.967
3482	89.8%	0.001	0.937
3484	88.3%	0.000	0.976
3498	85.0%	0.006	0.665
3507	90.9%	0.002	0.894
3519	72.4%	0.001	0.957
3521	90.0%	0.001	0.918
3523	83.1%	0.001	0.918
3527	85.9%	0.000	0.964
3530	59.4%	0.001	0.912
3533	11.1%	0.011	0.536
3538	87.3%	0.001	0.926
3551	79.2%	0.001	0.960
3552	8.7%	0.003	0.786
3553	50.9%	0.005	0.729
3554	51.9%	0.009	0.578

3573	80.8%	0.001	0.923
3579	74.8%	0.001	0.906
3583	83.6%	0.000	0.990
3587	91.3%	0.001	0.927
3591	84.2%	0.004	0.783
3593	85.5%	0.001	0.961
3594	84.3%	0.001	0.924
3595	89.6%	0.000	0.991
3596	92.7%	0.001	0.954
3597	89.5%	0.000	0.972
3598	94.4%	0.001	0.944
3600	70.0%	0.005	0.707
3601	91.0%	0.000	0.991
3602	86.8%	0.001	0.910
3603	87.2%	0.001	0.934
3604	91.7%	0.006	0.665
3605	88.4%	0.001	0.947
3606	72.2%	0.001	0.924
3607	90.2%	0.000	0.988
3608	67.8%	0.001	0.932
3609	84.0%	0.003	0.825
3625	73.5%	0.003	0.816
3626	91.4%	0.001	0.903
3633	75.6%	0.002	0.849
3639	83.5%	0.000	0.989
3658	80.0%	0.016	0.442
3659	84.9%	0.001	0.902
3687	64.8%	0.002	0.890
3702	85.1%	0.000	0.988
3728	100.0%	0.000	1.000
3769	88.6%	0.000	0.989
3794	77.6%	0.001	0.955
3826	42.0%	0.002	0.891
3847	77.8%	0.000	0.985
3862	63.6%	0.021	0.376
3879	77.5%	0.001	0.945
3904	89.9%	0.000	0.993
3932	69.0%	0.002	0.837
3942	93.9%	0.000	0.967

3959	78.9%	0.002	0.844
3969	100.0%	0.000	1.000
3972	92.4%	0.000	0.970
3973	82.3%	0.001	0.957
3975	94.3%	0.002	0.892
3976	86.7%	0.000	0.977
3977	71.4%	0.000	0.982
3978	78.3%	0.000	0.975
3979	80.2%	0.000	0.977
3980	82.8%	0.001	0.960
3981	84.5%	0.000	0.995
3982	90.9%	0.000	0.985
3983	88.4%	0.000	0.973
3984	70.4%	0.004	0.766
3985	95.7%	0.001	0.936
3998	79.8%	0.002	0.891
4014	95.5%	0.000	0.975
4018	82.2%	0.001	0.903
4035	90.6%	0.001	0.953
4039	75.5%	0.002	0.879
4068	65.0%	0.011	0.527
4088	83.3%	0.001	0.953
4122	85.2%	0.002	0.890
4126	90.0%	0.001	0.908
4220	93.9%	0.000	0.967
4221	71.4%	0.003	0.827
4235	61.0%	0.002	0.842
7685	89.4%	0.001	0.953
7718	65.0%	0.003	0.817
7722	75.0%	0.002	0.837
7728	69.8%	0.004	0.761
7734	93.3%	0.000	0.981
7738	88.1%	0.000	0.991
7758	89.3%	0.003	0.787
7765	80.2%	0.000	0.970
7766	87.9%	0.000	0.984
7772	78.4%	0.000	0.970
7780	50.0%	0.042	0.233
7782	100.0%	0.000	1.000

7785	84.5%	0.000	0.965
7802	83.3%	0.000	0.980
7807	65.1%	0.002	0.893
7821	74.6%	0.003	0.826
7833	71.1%	0.002	0.893
7834	81.3%	0.001	0.926
7845	75.8%	0.001	0.938
7857	0.0%	0.000	1.000
7878	60.9%	0.005	0.710
7885	86.1%	0.000	0.979
7886	81.4%	0.001	0.904
7888	90.1%	0.000	0.983
7892	77.5%	0.001	0.916
7910	100.0%	0.000	1.000
7911	85.7%	0.004	0.743
7913	78.9%	0.002	0.844
7919	79.3%	0.006	0.691
7920	80.8%	0.002	0.856
7929	92.6%	0.000	0.990
7931	89.0%	0.000	0.982
7942	97.7%	0.000	0.998
7955	87.8%	0.001	0.939
7964	55.8%	0.006	0.688
7985	75.8%	0.003	0.811
7997	88.0%	0.000	0.966
7998	76.5%	0.005	0.705
8000	84.6%	0.000	0.988
8005	88.4%	0.000	0.971
8018	99.0%	0.000	0.996
8027	76.6%	0.003	0.819
8029	81.6%	0.000	0.966
8030	71.1%	0.005	0.735
8063	79.6%	0.003	0.793
8067	82.7%	0.000	0.967
8079	79.2%	0.001	0.943
8102	91.5%	0.001	0.921
8111	68.8%	0.002	0.881
8119	30.6%	0.001	0.954
8129	71.5%	0.001	0.934

8130	83.5%	0.001	0.940
8131	88.5%	0.000	0.989
8132	88.2%	0.000	0.987
8133	87.3%	0.000	0.990
8134	88.2%	0.000	0.984
8135	92.2%	0.000	0.987
8136	87.3%	0.000	0.964
8142	79.1%	0.002	0.868
8143	91.7%	0.001	0.947
8149	0.0%	0.000	1.000
8160	84.6%	0.000	0.985
8163	80.4%	0.003	0.818
8166	61.6%	0.002	0.841
8167	26.7%	0.007	0.660
8180	86.8%	0.001	0.910
8181	89.0%	0.001	0.922
8199	81.6%	0.000	0.992
8228	90.6%	0.000	0.981
8229	95.2%	0.001	0.958
8242	83.0%	0.000	0.998
8260	66.7%	0.004	0.774
8261	88.4%	0.000	0.964
8262	79.7%	0.002	0.844
8263	89.5%	0.000	0.986
8265	91.3%	0.000	0.993
8277	89.4%	0.000	0.965
8282	55.0%	0.012	0.506
8284	84.0%	0.005	0.702
8288	100.0%	0.000	1.000
8294	75.7%	0.002	0.888
8295	89.2%	0.001	0.916
8300	50.0%	0.125	0.092
8301	33.3%	0.037	0.255
8302	71.4%	0.029	0.303
8305	100.0%	0.000	1.000
8308	100.0%	0.000	1.000
8313	90.8%	0.000	0.981
8320	90.5%	0.004	0.755
8323	57.6%	0.001	0.908

8330	100.0%	0.000	1.000
8332	87.5%	0.014	0.481
8334	50.0%	0.125	0.092
8355	85.5%	0.001	0.957
8369	86.4%	0.000	0.963
8373	80.0%	0.001	0.943
8387	82.0%	0.000	0.965
8396	80.9%	0.000	0.984
8397	80.0%	0.000	0.965
8399	95.5%	0.001	0.928
8401	85.1%	0.001	0.919
8403	80.7%	0.000	0.972
8405	87.1%	0.002	0.888
8407	69.2%	0.001	0.895
8411	88.4%	0.000	0.991
8412	100.0%	0.000	1.000
8414	91.7%	0.002	0.856
8415	100.0%	0.000	1.000
8419	85.5%	0.000	0.963
8421	78.9%	0.001	0.924
8425	81.4%	0.004	0.782
8426	77.6%	0.001	0.919
8427	57.1%	0.005	0.717
8430	72.2%	0.011	0.532
8432	100.0%	0.000	1.000
8438	16.7%	0.006	0.686
8441	53.6%	0.009	0.588
8504	54.5%	0.023	0.360
8506	86.4%	0.000	0.997
8507	85.8%	0.000	0.972
8508	91.3%	0.001	0.960
8510	88.4%	0.000	0.994
8511	79.2%	0.000	0.983
8512	83.7%	0.000	0.982
8513	72.3%	0.002	0.840
8537	61.5%	0.002	0.854
8538	81.3%	0.000	0.967
8542	80.7%	0.000	0.981
8546	89.4%	0.001	0.938

8550	81.6%	0.001	0.941
8551	92.3%	0.000	0.972
8553	96.7%	0.001	0.922
8559	93.0%	0.001	0.932
8561	86.5%	0.000	0.976
8563	90.1%	0.001	0.910
8566	78.9%	0.009	0.591
8568	79.5%	0.004	0.752
8570	88.9%	0.011	0.536
8571	82.4%	0.001	0.922
8573	69.4%	0.003	0.787
8575	92.6%	0.001	0.945
8577	63.6%	0.007	0.644
8579	66.7%	0.037	0.255
8580	73.4%	0.001	0.941
8598	74.3%	0.000	0.991
8608	43.5%	0.005	0.703
8611	67.6%	0.003	0.797
8618	88.0%	0.001	0.952
8624	83.7%	0.000	0.985
8626	84.6%	0.010	0.558
8632	58.2%	0.001	0.939
8635	70.6%	0.012	0.509
8636	100.0%	0.000	1.000
8638	100.0%	0.000	1.000
8639	93.9%	0.002	0.880
8640	100.0%	0.000	1.000
8641	100.0%	0.000	1.000
8644	100.0%	0.000	1.000
8645	75.0%	0.047	0.213
8650	92.1%	0.000	0.964
8651	79.1%	0.001	0.936
8653	50.0%	0.125	0.092

Overall reliability scores (i.e., median of provider-level reliability [R_median], minimum [R_min], maximum [R_max]) by year, and the overall variance between sites, are summarized below.

Overall reliability scores by year, 2010-2014

Year	% suppressed	Var_between	R_median	R_min	R_max

2010	61.8	0.051	0.983	0.290	1.000
2011	65.5	0.046	0.982	0.267	1.000
2012	69.9	0.038	0.979	0.338	1.000
2013	76.1	0.020	0.967	0.211	1.000
2014	80.3	0.013	0.954	0.092	1.000

Reliability scores varied across providers by year. The proportion of providers with reliability greater than or equal to 0.9, 0.8, and 0.7 are shown below.

Distribution of provider-level reliability scores by year, 2010-2014

Year	N	≥0.9 n (%)	≥0.8 n (%)	≥0.7 n (%)
2010	846	764 (90.3)	809 (95.6)	826 (97.6)
2011	811	721 (88.9)	766 (94.5)	786 (96.9)
2012	816	713 (87.4)	775 (95.0)	794 (97.3)
2013	823	657 (79.8)	738 (89.7)	772 (93.8)
2014	813	595 (73.2)	690 (84.9)	751 (92.4)

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

There is no established cut-off for minimum reliability level. Values above 0.7 are considered sufficient to see differences between providers and the mean, and values above 0.9 are considered sufficient to see differences between pairs of providers (RAND Corporation, TR-653-NCQA, 2009).

Each year, the majority of provider-level reliability scores were greater than 0.9, and more than 90% of providers had reliability scores of 0.7 or greater. Therefore, the reliability of viral suppression can be considered to be sufficient to identify real differences in performance across providers. As previously mentioned, sample size is another driver of reliability and likely contributed to the lowest reliability scores (e.g., in 2014 site 8645 had a reliability of 0.21, and reported 3 of 4 patients with a medical visit were virally suppressed). However, median reliability was consistently over 0.95 during 2010-2014 and can help to support the conclusion that the reliability of this measure can be considered very good.

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)*

5. 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

6. Face validity of the performance score was gained through a structured presentations (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*)

5. 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.
6. Sixty-nine individuals/organizations submitted 239 pieces of comments. Twenty 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?*

5. 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.
6. 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 (20) 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)

NA ☒ no exclusions — *skip to section 2b4*

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

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

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

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

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

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

☒ **No risk adjustment or stratification**

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

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

☐ **Other,** Click here to enter description

2b4.1.1 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.

The Ryan White HIV/AIDS Program provides a comprehensive system of care that includes primary medical care and essential support services for people living with HIV who are uninsured or underinsured. The Program works with cities, states, and local community-based organizations to provide HIV care and treatment services to more than half a million people each year. The Program reaches approximately 52% of all people diagnosed with HIV in the United States.

As indicated in data presented earlier, the Ryan White HIV/AIDS Program is a public health, safety net program providing care to a high proportion of racial/ethnic minority, transgender, unstable housing, and low income people living with HIV. Many of people served by the Ryan White HIV/AIDS Program represent sociodemographics factors incorporate in risk adjusting models by many measures stewards. As a result, the Ryan White HIV/AIDS Program does not adjust for risk in its performance measures. Rather, it is a fundamental aspect of the Ryan White HIV/AIDS Program to identify disparities and work to improve quality of care for subpopulations. Additionally, this measure is not used for pay-for-performance, bonuses, or penalties.

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 viral suppression across providers, by year. Performance scores were broken into the bottom 10% and top 90% providers to better characterize the gaps that remain across providers. Moreover, performance scores were examined with respect to NHAS 2020 Indicator 6: increase the percentage of persons with diagnosed HIV infection who are virally suppressed to at least 80 percent.

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*)

	% Patients with viral suppression across providers					Providers achieving ≥80% suppression		
Year	Mean	SD	Median	10th %ile	90th %ile	N	n	%
2010	60.6	23.8	67.8	19.5	82.8	846	145	17.1
2011	64.7	22.1	71.4	31.9	84.9	811	207	24.5
2012	69.9	20.3	75.6	40.2	88.0	816	277	32.7
2013	76.1	17	80.7	57.1	90.2	823	435	51.4
2014	80.3	15.5	84.2	65.0	93.1	813	530	65.2

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. In 2014, the bottom 10% of providers had viral suppression rates of 65.0% or lower; the top 90% of providers had viral suppression rates of 93.1% or higher. While this gap appears to be narrowing over time, a meaningful difference of 28.1 percentage points remains, demonstrating the value of the measure in identifying sites based on poor performance relative to the top performers.

Provider-level performance differences observed in the table above also underscore improvements in the proportion of patients with viral suppression in achieving 80% viral suppression. In 2014, of 813 providers, 530 (65.2%) had at least 80% of patients reach viral suppression. Additionally, the overall percentage of patients

with viral suppression was 80.3%; however, given the large population that the RWHAP serves, even the poorest performing sites (e.g., bottom 10%) represent a substantial number of patients.

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

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

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*)

Based on the method used to calculate the viral suppression performance score, conducting missing data analysis is not applicable for this measure. Specifically, the logic used to determine the number of patients with viral suppression (a) relied on available viral load data (i.e., count data in the form of copies/ml). If the results of a viral load test were present for a given patient, the patient was considered to have had a viral load test. If no viral load count data were present for a given patient, the patient was considered to have had no viral load test during the measurement year (b1). Additionally, because viral suppression was conditional on patients having available viral load count data, among patients with a viral load test, the viral suppression status could only be calculated for those with count data, so there were no missing data on viral suppression (<200 copies/mL).

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; <u>if no empirical sensitivity analysis</u>, identify the approaches for handling missing data that were considered and pros and cons of each*)

N/A (see 2b7.1)

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; <u>if no empirical analysis</u>, provide rationale for the selected approach for missing data*)

N/A (see 2b7.1)

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 collection and availability: The data used for testing and operational use of this measure are readily available within patient health records and provided annually to the Ryan White HIV/AIDS Program through the reporting of the Ryan White Service Report (approved by the Office of Management and Budget 0915-0323).

Missing data: A full analysis of missing data is provided in this submission.

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 Ryan White HIV/AIDS Program https://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio Public Health/Disease Surveillance Ryan White HIV/AIDS Program https://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio National HIV/AIDS Strategy https://www.aids.gov/federal-resources/national-hiv-aids-strategy/nhas-update.pdf Payment Program Medicaid Adult Core Set https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-core-set/index.html PQRS https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html?redirect=/pqri Regulatory and Accreditation Programs Ryan White HIV/AIDS Program https://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio 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 maintenance of endorsement), 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

Medicaid Adult Core Set
Sponsor: Federal government
Geographic area: Nationwide
Accountable entities: State Medicaid programs
Patients: Unknown

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

National HIV/AIDS Strategy
Sponsor: Federal government
Geographic area: Nationwide
Accountable entities: Federal agencies and service providers
Patients: All people living with HIV in the United States

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.
HIV viral suppression has been improving in the United States since the first release of publically available data. The Ryan White HIV/AIDS Program served more than 300,000 unduplicated patients annually between 2010-2014 across 2,000+ grant recipients and subrecipients. The Ryan White HIV/AIDS Program has experienced a 20-point increase in viral suppression from 61.8% in 2010 to 80.3% in 2014. Viral suppression has increased across all demographic groups and subpopulations.

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.**
N/A

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.
Starting in 2015, Health Resources and Services Administration began releasing December 1st – World AIDS Day – an annual data report (Ryan White HIV/AIDS Program Annual Client-Level Data Report) that contains data similar to those presenting in the report. Building upon the success of the state profiles (http://hab.hrsa.gov/stateprofiles/), Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived). A supplemental report exploring data for the eligible metropolitan areas and transitional grant areas and youth/young adults has been released as well as slides sets for fact sheets by program and population, special populations (http://hab.hrsa.gov/publications/hivaids-bureau-fact-sheets), and infographics (contained in fact sheets). Additionally, grant recipient level reports are prepared and disseminated to all Ryan White HIV/AIDS Program grant recipients.

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.**
Starting in 2015, Health Resources and Services Administration began releasing December 1st – World AIDS Day – an annual data report (Ryan White HIV/AIDS Program Annual Client-Level Data Report) that contains data similar to those presenting in the report. Building upon the success of the state profiles (http://hab.hrsa.gov/stateprofiles/), Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived). A supplemental report exploring data for the eligible metropolitan areas and transitional grant areas and youth/young adults has been released as well as slides sets for fact sheets by program and population, special populations (http://hab.hrsa.gov/publications/hivaids-bureau-fact-sheets), and infographics (contained in fact sheets). Additionally, grant recipient level reports are prepared and disseminated to all Ryan White HIV/AIDS Program grant recipients.

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 from Ryan White HIV/AIDS Program grant recipients and subrecipients regarding the feasibility and usefulness of the data presented in the Ryan White HIV/AIDS Program Annual Client-Level Data Report. Significant feedback has been provided about the timeliness and expansions of the data release. Grant recipient report using the data for benchmarking their program, setting goals/targets, and gaining a fuller understanding of all aspects of the Ryan White HIV/AIDS Program (i.e. other regions of the country). Grant recipients and subrecipients have also requested additional analyses. Health Resources and Services Administration responded with supplemental reports (Ryan White HIV/AIDS Program Supplemental Client-Level Data Report, Eligible Metropolitan Areas and Transitional Grant Areas; special population reports); slide decks for the overall client population and special populations; grant recipient reports; and infographics – all of which will be updated and released annually. Health Resources and Services Administration plans to release additional analyses and special reports this year based on feedback from Ryan White HIV/AIDS Program grant recipients and subrecipients.

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

4d2.3. **Summarize the feedback obtained from other users**
Ryan White HIV/AIDS Program national partners (national organizations that represent grant recipients, subrecipients, and patients) has provided antidotal feedback regarding the timeliness, feasibility, and usability of the release of the Ryan White HIV/AIDS Program Annual Client-Level Data Report, supplemental reports, slide decks, fact sheets, and infographics. The national partners encourage the continued release of the data in all its formats.

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.

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)**
0407 : HIV/AIDS: HIV RNA Control After Six Months of Potent Antiretroviral Therapy

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
2079 HIV Medical Visit Frequency
2080 Gap in HIV Medical Visits
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?
No

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.)
None

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.

No appendix **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

NATIONAL QUALITY FORUM

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 #: 3210
Measure Title: HIV viral load suppression
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 with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year.
Developer Rationale: Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. Antiretroviral therapy delays this progression and increases the length of survival.

Antiretroviral therapy 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). Viral suppression is a main goal of HIV treatment and an indicator of treatment success and reduction of potential HIV transmission. It is directly related to:

- Reduction in disease progression, incidence of opportunistic infections, the risk of both defining and non-AIDS- defining complications and the incidence and severity of chronic conditions.
- Reduction in the risk of transmitting HIV to a sexual or drug-using partner who does not have HIV.
- Improvement of immune function, quality of life, increase in time until development of AIDS increase in life expectancy.

Being virally suppressed is good for an HIV-positive person's overall health and preventing HIV infection from advancing to AIDS, the last stage of HIV infection.

- Durable viral suppression improves immune function and quality of life, prolongs life, lowers the risk of both AIDS-defining and non-AIDS-defining complications, and prolongs life. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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: Patients with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year. The outcome being measured is HIV viral suppression.

Denominator Statement: Patients, regardless of age, diagnosed with HIV during the first 3 months of the measurement year or prior to the measurement year who had at least one medical visit in the measurement year. The target population for this measure is all people living with HIV.
Denominator Exclusions: There are no patient exclusions.

Measure Type: Outcome
Data Source: Electronic Health Record (Only), Other
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

This measure is the new eMeasure version of the chart-abstracted measure #2082. The information provided for Evidence and Opportunity for Improvement is identical to that submitted for #2082. Measure #2082 will be discussed first – the ratings for evidence and opportunity for improvement will automatically be assigned to this eMeasure without further discussion.

Evidence Summary:

- 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.
- According to the developer, viral suppression is a main goal of HIV treatment and an indicator of treatment success and reduction of potential HIV transmission.
- Being virally suppressed is good for a HIV-positive person's overall health and preventing HIV infection from advancing to AIDS, the last stage of HIV infection. Viral suppression is directly related to:
    - Reduction in disease progression, incidence of opportunistic infections, the risk of both defining and non-AIDS- defining complications and the incidence and severity of chronic conditions.
    - Reduction in the risk of transmitting HIV to a sexual or drug-using partner who does not have HIV.
    - Improvement of immune function, quality of life, increase in time until development of AIDS and increase in life expectancy.
- The developer provided multiple guidelines for the administration of antiretroviral therapy and viral load monitoring intervals for adults, adolescents and pregnant women.
- The developer provided sufficient evidence demonstrating that antiretroviral therapy and viral suppression reduce morbidity and mortality associated with HIV.

Questions for the Committee:

- *Does the Committee agree that a viral load of less than 200 copies/mL leads to improved patient outcomes for patients with a diagnosis of HIV?*
- *Does the Committee agree that there is at least one thing that the provider can do to achieve a change in the viral load of patients diagnosed with HIV? If so, does the Committee agree there is no need for repeat discussion and vote on Evidence?*

Guidance from the Evidence Algorithm: Health outcome measure (Box 1)→The relationship between the outcome and at least one process is identified and supported by the stated rationale (Box 2)→Pass

Preliminary rating for evidence: ☒ Pass ☐ No Pass

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

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

- Per the developer, currently there is no performance data available from the eCQM. However, the developer provided 2014 nationwide data from the CDC that estimates that although 86% of people living with HIV have been diagnosed, only 30% have achieved viral suppression.
- The developer provided the following facility-level performance rates from the Ryan White HIV/AIDS Program Services Report (RSR) from 2010 – 2014 from the existing chart-abstracted measure, #2082:

	2014	2013	2012	2011	2010
Rate	80.8	76.1	69.9	65.5	61.8
Pts w/ ≥1 medical visit (den)	316,087	327,618	335,408	327,744	324,455
Pts w/viral suppression (num)	255,342	249,436	234,505	214,650	200,584
Mean	80.3	76.1	69.9	64.7	60.6
Median	84.2	80.7	75.6	71.4	67.8
Standard Deviation	15.5	17.0	20.3	22.1	23.8
10th percentile	65.0	57.1	40.2	31.9	19.5
90th percentile	93.1	90.2	88.0	84.9	82.8
Min, Max	0.0, 100.0	0.0, 100.0	0.0, 100.0	0.0, 100.0	0.0, 100.0
Pts w/viral load test performed	293,237 (92.8)	297,066 (90.7)	289,563 (86.3)	273,241 (83.4)	264,630 (81.6)
# of facilities	813	823	816	811	846

Disparities:

- The developer provided the following 2010 – 2014 viral suppression rates from the chart-abstracted measure, #2082:

Age	2014	2013	2012	2011	2010
<13	35.5	37.9	36.3	30.5	36.3
13-14	83.8	81.6	76.2	69.2	60.9
15-19	71.5	65.4	57.3	53.8	51.0
20-24	68.2	60.2	50.7	46.9	41.8
25-29	72.5	66.3	58.3	53.5	48.9
30-34	75.9	70.5	63.5	59.5	55.2
35-39	78.0	73.7	67.3	63.0	60.0

40-44	79.9	75.9	70.3	66.0	62.5
45-49	82.1	78.4	72.9	68.5	64.9
50-54	85.7	81.9	77.0	72.0	67.8
60-64	87.1	83.4	78.5	73.8	69.7
≥65	88.4	84.7	80.7	74.7	70.7
Race/Ethnicity					
American Indian/Alaska Native	82.4	74.2	72.3	68.2	64.9
Asian	83.4	78.5	72.4	67.6	64.8
Black/African American	77.4	72.3	66.0	61.2	56.9
Hispanic/Latino	82.6	78.2	72.7	67.6	62.8
Native Hawaiian/Pacific Islander	75.4	67.9	65.2	67.9	57.9
White	83.8	80.2	74.5	70.3	68.3
Multiple Races	83.6	78.0	71.7	66.0	66.8
Gender					
Male	80.7	76.1	70.2	65.7	62.3
Female	79.8	75.0	68.8	63.7	58.9
Transgender	73.6	72.1	66.3	60.1	55.4

- The developer also provided performance rates based on transmission risk, health care coverage, provider type and National HIV/AIDS Strategy (NHAS) populations from the existing chart-abstracted measure, #2082.

Questions for the Committee:

- *Without data from the eMeasure as specified, do you agree that there is a quality problem achieving viral suppression for patients diagnosed with HIV? Is there opportunity for improvement?*
- *Is a national performance measure warranted?*
- *Are you aware of evidence that disparities exist in this area of healthcare?*

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)

1a. Evidence

*Same as measure 2082

*This is an outcome measure up for new eMeasure review. HIV viral load suppression is linked with decreased disease progression, incidence of OIs, and other clinically relevant outcomes. Significant evidence supports the importance of viral load suppression.
I am not aware of any new studies that alters the evidence base.
However, the viral load indicated by the viral load measure is set at <200. With improvements in antiretroviral therapy and assays to measure viral load, guidelines support viral suppression which would now be considered levels much lower than 200 (i.e. less than 20 or undetectable). Current performance data suggests there is still opportunity for improvement even with a permissive cutoff of 200. The developer should consider reassessing the cutoff range in the future as rates of compliance increase."

1b. Performance Gap
*Same as measure 2082

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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): electronic health record (EHR). This is an eMeasure.

Specifications:

- HQMF specifications for the eMeasure are included in the document set on SharePoint. See eMeasure Technical Advisor review below.
- The level of analysis is at the facility-level.
- The numerator includes the number of patients in the denominator with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year.
- The denominator includes the number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the measurement year. To be included in the denominator, patients must meet all of the following conditions/events:
  - Patients of any age during the measurement year
  - Patients diagnosed with HIV during the first 3 months of the measurement year or prior to the measurement year
  - Patients who had at least one medical visit during the measurement year
- There are no patient exclusions.
- The value sets needed to calculate the numerator and denominator are included in the specifications.
- The calculation algorithm is included.

Questions for the Committee:

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

eMeasure Technical Advisor(s) review:

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	**N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM**
Value Sets	**The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC**
Measure logic is unambiguous	**Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously**

Feasibility Testing	**The submission contains a feasibility assessment that addresses data element feasibility and follow-up with measure developer indicates that the measure logic is feasible based on assessment by EHR vendors**

2a2. Reliability Testing **Testing attachment**

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.

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 used for testing included 34 synthetic patients created in the Bonnie testing system simulating the year 2012. The developer tested the following data elements using the Bonnie testing tool to evaluate the measure logic:
  - Patient name
  - Date of birth
  - Race
  - Ethnicity
  - Gender
  - Payer
  - Diagnosis
  - Laboratory tests and associated results
  - Encounters
- The patient bundle's demographics were designed to mimic the HIV/AIDS population, specifically drawing from the patient characteristics collected via the Ryan White HIV/AIDS Program Services Report (RSR).
- Data element validity testing was performed and will count for data element reliability – see validity testing section.
- The developer provided reliability results from the chart-abstracted measure (#2082) and stated, "Currently, there is no performance data available to test the eCQM. However, the chart-abstracted version of this measure has been in use in national quality reporting programs since as early as 2010."

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results from the Bonnie tool demonstrate sufficient reliability so that differences in performance can be identified?*
- *Do you agree that the reliability results of the eMeasure will be comparable to the chart-abstracted measure (#2082)?*

Guidance from the Reliability Algorithm: Precise specifications (Box 1) → Empirical reliability testing (Box 2) → Empirical validity testing of patient-level data (Box 3) → Refer to validity testing of patient-level data elements using Bonnie tool (Box 10) → Method appropriate for legacy eMeasures (Box 11) → Moderate (highest eligible rating is MODERATE)

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

2b. Validity

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:

- *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.

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:

- The Bonnie testing environment, with 34 synthetic patient records, were used to test the measure logic and data elements.
  - For each synthetic patient, an expected result was assigned to reflect an expected result of the measure. The synthetic patients were then run against the HQMF output loaded into Bonnie, which "calculates" a measure result for each patient and evaluates it against the expected result.
  - A patient is considered to pass Bonnie testing when the expected result matches the "calculated" result.
- The developer conducted the following testing on synthetic patients:
  - 100% logic coverage: The bundle of synthetic patients collectively includes all data elements and conditions that are specified within the measure logic.
  - Edge case testing: Data elements that test the upper or lower boundary of measure logic conditions.
  - Negative testing: Use of test cases that do not evaluate positively against the measure logic but are otherwise clinically relevant and realistic.
- The developer used references cited within the chart abstracted measure specifications to ensure the eCQM logic maintained alignment with the clinical intent of the chart abstracted measure.
- In addition to Bonnie testing, the measure specifications were reviewed independently by three eCQM experts to confirm the logic was syntactically correct, using appropriate and current versions of the eCQM standards and terminologies, and consistent with the intent of the chart-abstracted measure.

Validity testing results: See Bonnie testing results

- The testing results from the Bonnie tool reached 100% coverage and confirmed there was a test case for each pathway of logic (negative and positive test cases).
- The measure also had a 100% passing rate which confirmed that all the test cases performed as expected.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results from the Bonnie tool demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the reliability results of the eMeasure will be comparable to the chart-abstracted measure (#2082)?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

N/A

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

- The Ryan White HIV/AIDS Program is a public health, safety net program providing care to a high proportion of racial/ethnic minority, transgender, unstable housing, and low income people living with HIV. Many of people served by the Ryan White HIV/AIDS Program represent sociodemographics factors incorporated in risk adjusting models by many measures stewards. As a result, the Ryan White HIV/AIDS Program does not adjust for risk in its performance measures. Rather, it is a fundamental aspect of the Ryan White HIV/AIDS Program to identify disparities and work to improve quality of care for subpopulations. Additionally, this measure is not used for pay-for-performance, bonuses, or penalties.

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

- The developer provided the percentage of patients with viral suppression across providers from the chart-abstracted measure.

Question for the Committee:

- *Does the Committee agree the e-Measure will demonstrate similar results to the chart-abstracted measure?*

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- Per the developer, "The HQMF standard specifies that if data are unknown or missing, they shall fail the criterion. This constraint embodies the notion that absence of evidence is evidence of absence, i.e. data not present in a structured field from which the measure draws will not be considered for measure calculation. In certain cases, missing data may have no impact on the measure outcome for a given patient. For example, a data element used in a series of OR statements will not impact the measure outcome if another data element in the OR statement is present and meets all other defined constraints."
- All Bonnie synthetic patients with missing data performed according to the HQMF standard specification and as expected.

Guidance from the Validity Algorithm: Specifications consistent with evidence (Box 1) → Threats to validity assessed (Box 2) → Empirical validity testing (Box 3) → Empirical validity testing of data elements and measure logic using Bonnie tool (Box 10) → Method appropriate for legacy eMeasures (Box 11) → Moderate (highest eligible rates is MODERATE)

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

Committee pre-evaluation comments

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

2a1. Reliability Specifications

*Data elements are clearly defined. The eMeasure technical review is positive. It is stated that the level of analysis is at the facility level.

*This measure applies at the facility level. Numerator and denominator statements are clearly defined. There are no exclusions. the calculation algorithm is a simple rate/proportion without any risk adjustment.

I have no concerns that this measure would be inconsistently applied based on the provided definitions. "

*see 2082

2a2. Reliability Testing

*Data is provided on the reliability of data obtained from 34 simluated patients. Agree with the preliminary rating for reliability.

*The Bonnie testing system was used to simulate the data elements including dob, race, ethnicity, gender, payer, diagnosis, lab tests, encounters. This data set included 34 synthetic patitients. Although a chart abstracted measure #2082 has existed for several years, corresponding data for an eCQM has not existed. Reliability testing has not been performed with real data.

*Not clear that the test sample is adequate to generalize for widespread implementation
Results from the Bonnie tool are from only a few synthetic patients so may not be of sufficient reliability

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

*the specifications are consistent with the evidence

2b2. Validity Testing
*The proposed eMeasure performed well on the sample of simulated patients.

*Validity testing was performed in the bonnie testing environment using 34 synthetic patients. It passed 100% logic coverage, edge case and negative testing.
The eCQM logic was reviewed by 3 independent ECQM experts.
Although 34 synthetic patients seems like a small number of patients, there was a test case for each logic pathway. "

*Not clear that the test sample is adequate to generalize for widespread implementation
The results from the Bonnie tool demonstrate sufficient validity so that conclusions about quality can be made
I agree that the reliability results of the eMeasure will be comparable to the chart-abstracted measure (#2082)"

2b3-7. Threats to Validity
*A similar question to measure 2082: It is stated that the level of analysis is at the facility level, it is somewhat unclear whether this measure is intended to be a measure of quality for individual providers or for facilities.

*There are no exclusions nor risk adjustment as the developer provides care to a diverse patient population.
As with the chart abstracted measure, this measure shows significant divergence in rates across providers. This data though is presented on the provider level, as calculated for the chart abstracted measure,. This eMeasure is testing at the facility level.
The developer has criteria for dealing with missing data. If data are unknown or missing, they shall fail the criterion.
Some missing data may not impact compliance with the measure when in a series of OR statements. Missing data was in Bonnie synthetic patient testing and performed as expected. "

*2b.5 the e-Measure should demonstrate similar results to the chart-abstracted measure

Criterion 3. Feasibility

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 provided information on the feasibility testing in the eMeasure Feasibility Scorecard. The developer did not identify specific EHRs used to test feasibility in ambulatory care. Instead, the developer stated that the feasibility assessment "conducted by consensus of a panel of MITRE clinical informatics, measure development, and eCQM experts."
- The developer provided a summary of the latest publicly available data on Meaningful Use EHR capabilities and provider performance on objectives and measures directly relevant for the eCQM's data elements:
    - CPOE – Meds
    - CPOE – Labs
    - Demographics
    - Lab test results
    - Problem list

- On a scale from 1 to 3 (3 = highest score), all of the data elements received a score of '3' except, "Encounter, Performed: Face-to-Face Interaction (2)" and "Patient Characteristic Payer (2)".
- Score 2 definition for data standards: Terminology standards for this data element are currently available, but it is not consistently coded to standard terminology in the EHR, or the EHR does not easily allow such coding.
    - For the Encounter, Performed: Face-to-Face Interaction data element, the developer stated that the Health IT Standards Committee recommends SNOMED-CT as the standard terminology for encounters, however SNOMED-CT isn't currently widely used for this purpose. The eCQM alllows for capture of encounters in SNOMED-CT as well as CPT, a more widely used terminology.
    - For the Patient Characteristic Payer, the developer stated that the Public Health Data Standards Consortium developed a standard to encode payer data that is increasingly being adopted. It is anticipated that this data element will be coded in a nationally accepted terminology standard (future state) because the 2015 Edition Health IT certification criteria requires the use of the source payment typology standard.
- Overall, the measure is currently 98.89% feasible and 99.44% feasible in 1 to 2 years according to the scorecard.
- The measure specifications include CPT® codes (requires a license to use) and SNOMED Clinical Terms® (requires a Unified Medical Language System (UMLS) license available for free from the National Library of Medicine).

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*
- *Does the feasibility scorecard demonstrate that the eMeasure can be implemented and feasibility concerns can be adequately addressed.*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility
*Agree with the preliminary rating of high given the data provided.

*Data elements for this eMeasure are generated during routine care, although the developer did not identify specific EHRs used to test feasibility. Instead • the developer stated that the feasibility assessment "conducted by consensus of a panel of MITRE clinical informatics, measure development, and eCQM experts."
While there are standards formats for all data elements for this emeasure, utilization of EHRs that can electronically capture this data is a potential challenge if EHRs are not widely used. Meaningful use data shows some variation in % providers reporting data per the feasibility scorecard with, if interpreted correctly, some very low rates for reporting - e.g. Stage 1 MU providers in 2015 program year which only required use of EHR to order labs (not capture lab results). It appears that the number of providers that get lab results increases year over year but this is a potential threat to the measure.

*The required data elements are routinely generated and used during care delivery
The eMeasure Feasibility Score Card demonstrates acceptable feasibility in multiple EHR systems and sites"

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**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- This newly developed eMeasure is not currently in an accountability program; however, it was reviewed by NQF's Measure Applications Partnership (MAP) for consideration in CMS' Merit Based Incentive Payment System (MIPS). See MAP feedback below.

Improvement results:

- Based on the chart-abstracted measure, the Ryan White HIV/AIDS Program has experienced a 20-point increase in viral suppression from 61.8% in 2010 to 80.3% in 2014. Viral suppression has increased across all demographic groups and subpopulations.

Unexpected findings (positive or negative) during implementation:

- The developer did not provide any unexpected findings during implementation.

Vetting of the measure:

- Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived).
- Ryan White HIV/AIDS Program national partners (national organizations that represent grant recipients, subrecipients, and patients) has provided antidotal feedback regarding the timeliness, feasibility, and usability of the release of the Ryan White HIV/AIDS Program Annual Client-Level Data Report, supplemental reports, slide decks, fact sheets, and infographics. The national partners encourage the continued release of the data in all its formats.
- During the initial development of the chart-abstracted 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 measures are reviewed for clinical relevance, change in scientific acceptability, and consistency with guidelines. The chart-abstracted measure has not been modified as a result of the annual reviews.

Feedback:

- The MAP agreed that this outcome measure addresses an important clinical area. However, it has not been fully tested as an e-CQM. Additionally, the performance data is in the process of being updated from the 2011 data. The measure would address an important issue regarding HIV viral suppression and would provide an additional mechanism for submitting data on this topic. MAP discussed the importance of this measure as it adds an additional outcome measure to the [MIPS] program. The MAP recommended supported this eCQM for rulemaking with the condition that it completes successful testing and the NQF Behavioral Health Standing Committee reviews the performance data to ensure a gap in care continues to exist.

Questions for the Committee:

- *How can the performance results from the eCQM measure be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*
- *How has the eCQM been vetted in real-world settings by those being measured 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 chart abstraction version of this measure (2082) has been in use and has demonstrated improvement over the past years.

*As a new measure, it is not currently being used in any accountability programs; however the chart review based companion measure is.
HRSA releases the data report in the same year as collected. It is sharedpublically. "

Criterion 5: Related and Competing Measures

Related or competing measures

- 0407 HIV/AIDS: HIV RNA Control After Six Months of Potent Antiretroviral Therapy (NCQA)
- 0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis (NCQA)
- 0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis (NCQA)
- 2079 HIV Medical Visit Frequency (HRSA)
- 2080 Gap in HIV Medical Visits (HRSA)
- 2083 Prescription of HIV Antiretroviral Therapy (HRSA)
- 3211 Prescription of HIV Antiretroviral Therapy (HRSA)
- 3210 HIV Viral Suppression (HRSA)
- 3010 HIV Medical Visit Frequency (HRSA)

Harmonization

- Per developer, harmonized with all measures except #0405 and #0409. Plans to harmonize with #0405 and #0409.

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 candidate 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**

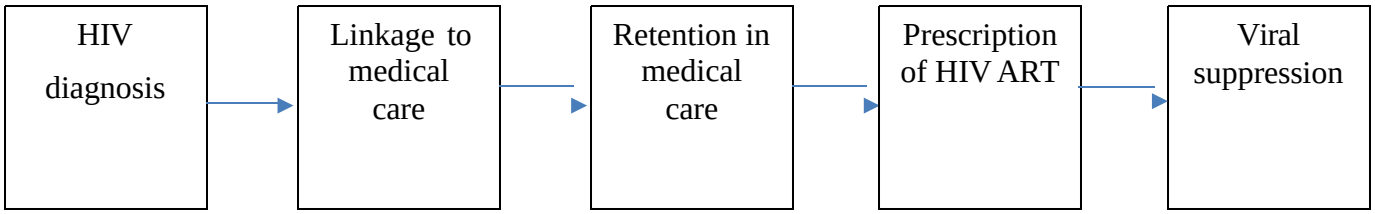
RATIONALE IF NOT ELIGIBLE: The measure is not eligible for Endorsement+ because empirical reliability and validity testing of the measure score was not conducted and the measure has not been vetted in real world settings by those being measured and other users.

Pre-meeting public and member comments

-

Measure Title: Viral Load Supression

1a.12 LOGIC MODEL



Althought the above diagram outlines the sequenctial septs of medical care that people living with HIV go through from initial diagnosis to achieving the goal of viral suppression. For some patients, this is a linear path with sustained viral suppression for many years. For other patients, there many 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 patients 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)

Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. Antiretroviral therapy delays this progression and increases the length of survival.

Antiretroviral therapy 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). Viral suppression is a main goal of HIV treatment and an indicator of treatment success and reduction of potential HIV transmission. It is directly related to:

- Reduction in disease progression, incidence of opportunistic infections, the risk of both defining and non-AIDS- defining complications and the incidence and severity of chronic conditions.
- Reduction in the risk of transmitting HIV to a sexual or drug-using partner who does not have HIV.
- Improvement of immune function, quality of life, increase in time until development of AIDS increase in life expectancy. Being virally suppressed is good for an HIV-positive person's overall health and preventing HIV infection from advancing to AIDS, the last stage of HIV infection.
- Durable viral suppression improves immune function and quality of life, prolongs life, lowers the risk of both AIDS-defining and non-AIDS-defining complications, and prolongs life. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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

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, Accessed November 18, 2016

https://aidsinfo.nih.gov/guidelines/html/1/adult-and-adolescent-arv-guidelines/0

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

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

Günthard HF, Saag MS, Benson CA, del Rio C, Eron JJ, Gallant JE, Hoy JF, Mugavero MJ, Sax PE, Thompson MA, Gandhi RT, Landovitz RJ, Smith DM, Jacobsen DM, Volberding PA. Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults 2016 Recommendations of the International Antiviral Society (IAS)–USA Panel. JAMA. 2016. https://www.iasusa.org/content/antiretroviral-drugs-treatment-and-prevention-hiv-infection-adults-2016-recommendations

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

Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents

Panel's Recommendations for Initiating Antiretroviral Therapy in Treatment-Naive Patients (pE1)

Antiretroviral therapy (ART) is recommended for all HIV-infected individuals, regardless of CD4 T lymphocyte cell count, to reduce the morbidity and mortality associated with HIV infection (**AI**).

Panel's Recommendations for Acute and Recent (Early) HIV Infection (pI1)

- Antiretroviral therapy (ART) is recommended for all individuals with HIV-1 infection **(AI)** including those with early HIV-1 infection.
- Once initiated, the goal of ART is to suppress plasma HIV-1 RNA to undetectable levels **(AIII)**. Testing for plasma HIV-1 RNA levels, CD4 T lymphocyte counts, and toxicity monitoring should be performed as recommended for patients with chronic HIV-1 infection **(AII)**.

Panel's Recommendations Regarding Virologic Failure of the Treatment-Experienced Patient (pH1)

- The goal of treatment for ART-experienced patients with drug resistance who are experiencing virologic failure is to establish virologic suppression (i.e., HIV RNA below the lower limits of detection of currently used assays) **(AI)**.
- Discontinuing or briefly interrupting therapy may lead to a rapid increase in HIV RNA and a decrease in CD4 cell count and increases the risk of clinical progression. Therefore, this strategy is **not** recommended in the setting of virologic failure **(AI)**.

Laboratory Testing, Plasma HIV-1 RNA (Viral Load) and CD4 Count Monitoring (pC5)

- Viral load is the most important indicator of initial and sustained response to ART (AI) and should be measured in all HIV-infected patients at entry into care (AIII), at initiation of therapy (AIII), and on a regular basis thereafter. For those patients who choose to delay therapy, repeat viral load testing while not on ART is optional (CIII).
- Plasma viral load should be measured before initiation of ART and within 2 to 4 weeks but no later than 8 weeks after treatment initiation or modification (AIII). The purpose of the measurements is to confirm an adequate initial virologic response to ART, indicating appropriate regimen selection and patient adherence to therapy. Repeat viral load measurement should be performed at 4- to 8-week intervals until the level falls below the assay's limit of detection (BIII).
- In virologically suppressed patients in whom ART was modified because of drug toxicity or for regimen simplification. Viral load measurement should be performed within 4 to 8 weeks after changing therapy (AIII). The purpose of viral load monitoring at this point is to confirm the effectiveness of the new regimen.
- 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).

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

https://aidsinfo.nih.gov/guidelines/html/3/perinatal-guidelines/0

Panel's Recommendations for HIV-Infected Pregnant Women Who Have Never Received Antiretroviral Drugs (pC27)

- All HIV-infected pregnant women should receive combination antiretroviral therapy (cART) to reduce the risk of perinatal transmission of HIV **(AI)**. The choice of regimen should take into account current adult treatment guidelines, what is known about the use of specific drugs in pregnancy, and the risk of teratogenicity (see Table 6 and Table 7).
- Consideration should be given to initiating cART as soon as HIV is diagnosed during pregnancy; earlier viral suppression is associated with lower risk of transmission. This decision may be influenced by CD4 T lymphocyte count, HIV RNA levels, and maternal conditions (e.g., nausea and vomiting) **(AIII)**. The benefits of early cART must be weighed against potential fetal effects of drug exposure.

Panel's Recommendations Regarding Lack of Viral Suppression During Antepartum Care (pC48)

- Because maternal antenatal viral load correlates with risk of perinatal transmission of HIV, suppression of HIV RNA to undetectable levels should be achieved as rapidly as possible **(AII)**.
- If an ultrasensitive HIV RNA assay indicates failure of viral suppression (after an adequate period of treatment):

- Assess adherence and resistance (if HIV RNA level is high enough for resistance testing) **(AII)**.
- Consult an HIV treatment expert and consider possible antiretroviral regimen modification **(AIII)**.
- Scheduled cesarean delivery is recommended for HIV-infected pregnant women who have HIV RNA levels >1,000 copies/mL near the time of delivery **(AII)**.

International Advisory Panel on HIV Care Continuum Optimization

- Where possible, jurisdictions should consider longitudinal cohort measurement of HIV service utilization and treatment outcomes to identify the means to maximize viral suppression through ensuring early access to ART and retention in care. (A IV) (p4)

World Health Organization:

- ARV drugs play a key role in HIV prevention. People taking ART who achieve optimal viral suppression are extremely unlikely to pass HIV to sexual partners. ARV drugs taken by people without HIV as PrEP or PEP are highly effective in preventing HIV acquisition. (p64)
- People starting treatment and carergivers should be informed that the first ART regimen offers the best opportunity for effective viral suppression, immune recovery and consequently clinical benefit and that successful ART requires all medications to be taken as prescribed. (p72)
- Access to ART should be the first priority for all age groups, and lack of testing for monitoring treatment response should not be a barrier to initiating ART. If viral load testing capacity is limited, it should be introduced in a phased approach. Examples of phased approaches include: …
    - using viral load initially as a targeted test to confirm treatment failure;
    - prioritizing viral load testing for pregnant and breastfeeding women, especially around the time of delivery, as sustained viral suppression is critical to prevention of transmission to the child, and documented high viral load at delivery is an indication for enhanced infant prophylaxis; (p134)

Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults, 2016 Recommendations of the International Antiviral Society–USA Panel

- HIV RNA level should be monitored every 4 to 6 weeks after treatment is initiated or changed until virus is undetectable (evidence rating AIa). (Box 5)
- After viral suppression is achieved, HIV RNA should be monitored every 3 months until suppressed for 1 year and at least every 6 months thereafter for adherent patients who remain clinically stable (evidence rating AIII). (Box 5)
- When virus has been suppressed for at least 2 years and CD4 cell count is persistently above 500/μL, repeat monitoring of CD4 cell count is not recommended unless virologic failure (evidence rating AIIa) or intercurrent immunosuppressive conditions occur or immunosuppressive treatments are initiated (evidence rating AIII). (Box 5)
- If the HIV RNA level remains above the limit of quantification by 24 weeks after starting new treatment or if rebound above 50 copies/mL occurs at any time, the assay should be repeated within 4 weeks to exclude impending virologic failure (evidence rating AIIa). (Box 5)
- For patients with persistent quantifiable HIV RNA between 50 and 200 copies/mL, reassessment for causes of virologic failure, evaluation again within 4 weeks, and close monitoring are recommended (evidence rating BIII). (Box 5)

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.

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

Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults, 2016 Recommendations of the International Antiviral Society–USA Panel

Table 1. Strength of Recommendation and Quality of Evidence Rating Scale

Rating	Definition
Strength of recommendation	
A	Strong support for the recommendation
B	Moderate support for the recommendation
C	Limited support for the recommendation

Quality of evidence	
Ia	Evidence for ≥1 randomized clinical trials published in the peer-reviewed literature
Ib	Evidence for ≥1 randomized clinical trials presented in abstract form at peer-reviewed scientific meetings
IIa	Evidence from nonrandomized clinical trials or cohorts or case-control studies published in the peer-reviewed literature
IIb	Evidence from nonrandomized clinical trials or cohorts or case-control studies published in the peer-reviewed scientific meeting
III	Recommendation based on panel's analysis of the accumulated available evidnce

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
VS_evidence-636177547737712934.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.
Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. Antiretroviral therapy delays this progression and increases the length of survival.

Antiretroviral therapy 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). Viral suppression is a main goal of HIV treatment and an indicator of treatment success and reduction of potential HIV transmission. It is directly related to:

- Reduction in disease progression, incidence of opportunistic infections, the risk of both defining and non-AIDS- defining complications and the incidence and severity of chronic conditions.
- Reduction in the risk of transmitting HIV to a sexual or drug-using partner who does not have HIV.
- Improvement of immune function, quality of life, increase in time until development of AIDS increase in life expectancy.

Being virally suppressed is good for an HIV-positive person's overall health and preventing HIV infection from advancing to AIDS, the last stage of HIV infection.

- Durable viral suppression improves immune function and quality of life, prolongs life, lowers the risk of both AIDS-defining and non-AIDS-defining complications, and prolongs life. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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.*
See attachment "VS 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 "VS 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.)*
There is no measure-specific web page for the electronic version of this measure.

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 an eMeasure **Attachment:** NQFXXX_HIVViralSuppression_Artifacts-636178423251224574.zip,NQFXXX_HIVViralSuppression_MeasureSubmissionForm-636178426319023569.docx

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:** HIVVLS_v4_6_Thu_Dec_15_20.35.00_CST_2016-636178423774443650.xls

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.
Not applicable

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).
Patients with a HIV viral load less than 200 copies/mL at last HIV viral load test during the measurement year. The outcome being measured is HIV viral suppression.

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).
The viral load suppression laboratory test is represented by the QDM element "Laboratory Test, Performed: HIV Viral Load" using "HIV Viral Load Grouping Value Set (2.16.840.1.113883.3.464.1003.120.12.1002)". The result of the laboratory test is modeled as an attribute of the Viral Load Suppression QDM element and represented as a numerical result associated with copies/mL as the reporting unit.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients, regardless of age, diagnosed with HIV during the first 3 months of the measurement year or prior to the measurement year who had at least one medical visit in the measurement year. The target population for this measure is all people living with HIV.

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).
The patient's HIV diagnosis is represented by the QDM element "Diagnosis: HIV" using "HIV Grouping Value Set (2.16.840.1.113883.3.464.1003.120.12.1003)". The patient's medical visits are represented by the following QDM elements:
- "Encounter, Performed: Face-to-Face Interaction" using "Face-to-Face Interaction Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1048)"
- "Encounter, Performed: Office Visit" using "Office Visit Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1001)"
- "Encounter, Performed: Outpatient Consultation" using "Outpatient Consultation Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1008)"
- "Encounter, Performed: Preventive Care - Established Office Visit, 0 to 17" using "Preventive Care - Established Office Visit, 0 to 17 Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1024)"
- "Encounter, Performed: Preventive Care Services - Established Office Visit, 18 and Up" using "Preventive Care Services - Established Office Visit, 18 and Up Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1025)"
- "Encounter, Performed: Preventive Care Services-Initial Office Visit, 18 and Up" using "Preventive Care Services-Initial Office Visit, 18 and Up Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1023)"
- "Encounter, Performed: Preventive Care- Initial Office Visit, 0 to 17" using "Preventive Care- Initial Office Visit, 0 to 17 Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1022)"

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
There are no patient exclusions.

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.)*
There are no patient exclusions.

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 patients who meet the initial population criteria as defined by eCQM logic;
2. Identify and count subset of the initial population that meet denominator criteria as defined by eCQM logic;
3. Identify and count subset of patients in the denominator that meet numerator criteria as defined by eCQM logic.
4. Calculate the performance measure rate: by dividing the number of patients in the numerator population by the number of patients in the denominator population.

Note: the eCQM logic criteria for each population is defined in a computable format in the eCQM specifications provided as an attachment to this submission.

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.
Electronic Health Record (Only), Other

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.
Not applicable.

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
NQFXXX_ViralSuppression_BonnieTestingAttachment-636177547742392964.zip,NQFXXX_HIVViralSuppression_MeasureTestingAttatchment.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.)
Yes

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*)**:** Click here to enter NQF number
Measure Title: HIV Viral Suppression
Date of Submission: 12/16/2016
Type of Measure:

☒ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***

☐Intermediate Clinical Outcome	☐ Cost/resource
☐ Process	☐ Efficiency
☐ Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Synthetic Bonnie test patients	☒ other: Synthetic Bonnie test patients

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*).

This measure is a legacy electronic clinical quality measure (eCQM) – an NQF endorsed measure that has been respecified into eMeasures and are currently used in federal quality programs. Per NQF modified testing requirements for legacy eCQMs, the measure was tested in the Bonnie testing tool. Bonnie is designed to validate eCQM specifications (HQMF output and value sets) against the measure's expected behavior for user-developed synthetic test patients.

The synthetic patient bundle used to test this measure was designed to simulate clinically relevant, realistic patient scenarios aligned with the target population for this measure. Full details on the Bonnie synthetic patient bundle used to test this measure are included in the Bonnie testing attachment.

For more information on Bonnie, please visit https://bonnie.healthit.gov/.

1.3. What are the dates of the data used in testing? The Bonnie test environment simulates the year 2012 as the measurement period.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice

☒ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☒ other: Synthetic Bonnie test patients

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*)

Not applicable. The Bonnie synthetic patient bundle was used to test the measure.

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*)

A test bundle of 34 patients was designed and built within the Bonnie testing tool to evaluate the measure logic. Information documented for each patient within the bundle include:

- Patient name
- Date of birth
- Race
- Ethnicity
- Gender
- Payer

Additional elements contained within the patient profiles as appropriate for testing against expected outcomes include:

- Diagnosis
- Laboratory tests and associated results
- Encounters

The patient bundle's demographics were designed to mimic the HIV/AIDS population, specifically drawing from the patient characteristics collected via the Ryan White HIV/AIDS Program Services Report (RSR).

The breakdown of test bundle demographics for the 34 patients included (represented by number of patients/percentage of bundle): males 23/68%; females 11/32%; American Indian/Alaska Native 1/3%; Asian 1/3%; Black/African American 15/44%; Native Hawaiian/Pacific Islander 0/0%; White 9/26%; Hispanic/Latino 8/24%; younger than 13 1/3%; 13-17 years old 1/3%; 18-24 years old 2/6%; 25-34 years old 6/18%; 35-44 years old 6/18%; 45-54 years old 10/29%; 55-65 years old 6/18%; older than 65 2/6%.
Full details on the Bonnie synthetic patient bundle used to test this measure, including human-readable and QRDA Category 1 format documents for each synthetic patient record, are included in the Bonnie testing attachment.

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.

The Bonnie patient test deck was used to satisfy all testing requirements for this measure. The testing results are further supported by testing data for the chart-abstracted version of this measure collected through the Health Resources and Services Administration HIV/AIDs Bureau's Ryan White HIV/AIDS Program Services Report.

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).

Patient sociodemographic variables considered in the analysis of the chart-abstracted version of this measure were included in the eCQM specifications and modeled in the Bonnie patient bundle. These variables included age, race, ethnicity, gender and payer.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

Currently, there is no performance data available to test the eCQM. However, the chart-abstracted version of this measure has been in use in national quality reporting programs since as early as 2010.

The most recent reliability analysis of the chart-abstracted measure was confirmed 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.

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*)

Overall reliability scores (i.e., median of provider-level reliability [R_median], minimum [R_min], maximum [R_max]) by year, and the overall variance between sites, are summarized below.

Overall reliability scores by year, 2010-2014

Year	% suppressed	Var_between	R_median	R_min	R_max
2010	60.6	0.051	0.983	0.290	1.000
2011	64.7	0.046	0.982	0.267	1.000
2012	69.9	0.038	0.979	0.338	1.000
2013	76.1	0.020	0.967	0.211	1.000
2014	80.3	0.013	0.954	0.092	1.000

Reliability scores varied across providers by year. The proportion of providers with reliability greater than or equal to 0.9, 0.8, and 0.7 are shown below.

Distribution of provider-level reliability scores by year, 2010-2014

Year	N	≥0.9 n (%)	≥0.8 n (%)	≥0.7 n (%)
2010	846	764 (90.3)	809 (95.6)	826 (97.6)
2011	811	721 (88.9)	766 (94.5)	786 (96.9)
2012	816	713 (87.4)	775 (95.0)	794 (97.3)
2013	823	657 (79.8)	738 (89.7)	772 (93.8)
2014	813	595 (73.2)	690 (84.9)	751 (92.4)

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

There is no established cut-off for minimum reliability level. Values above 0.7 are considered sufficient to see differences between providers and the mean, and values above 0.9 are considered sufficient to see differences between pairs of providers (RAND Corporation, TR-653-NCQA, 2009).

Each year, the majority of provider-level reliability scores were greater than 0.9, and more than 90% of providers had reliability scores of 0.7 or greater. Therefore, the reliability of viral suppression can be considered to be sufficient to identify real differences in performance across providers. As previously mentioned, sample size is another driver of reliability and likely contributed to the lowest reliability scores (e.g., in 2014 site 8645 had a reliability of 0.21, and reported 3 of 4 patients with a medical visit were virally suppressed). However, median reliability was consistently over 0.95 during 2010-2014 and can help to support the conclusion that the reliability of this measure can be considered very good.

2b2. VALIDITY TESTING

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

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

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

The Bonnie testing environment was used to test the validity of the measure logic and data elements. For each Bonnie synthetic patient, an expected measure result was assigned to reflect the expected outcome of the measure given the specific patient scenario and associated data. The synthetic patients were run against the HQMF output loaded into Bonnie, which produces a measure outcome for each patient and evaluates it against the expected outcome. A patient is considered to pass Bonnie testing when the expected outcome matches the actual outcome, e.g. when a patient is expected to be in the numerator population and the computation of the synthetic patient data against the eCQM logic places the patient in the numerator.

In order to achieve a rigorous, clinically relevant test bundle, synthetic patients were designed following the below principles and test areas:

- Clinical relevance. References cited within the chart abstracted measure specification were used to design clinically relevant, realistic patient profiles for the measure's target population. This approach ensured the eCQM logic maintained alignment with the clinical intent of the chart abstracted measure.
- 100% logic coverage: The resulting bundle of synthetic patients collectively includes all data elements and conditions logic that are specified within the measure logic, including at least one patient evaluating against each measure population pathway. Fully testing the measure logic increases test rigor and mitigates risk of unexpected outcomes.
- Edge case testing. Edge cases refer to those data elements that test the upper or lower boundary of measure logic conditions, e.g. a diagnosis starting on the latest qualifying date or an HIV viral load result equal to the highest qualifying value. Edge cases are designed to test each edge that exists within each measure population.
- Negative testing. Negative testing involves use of test cases do not evaluate positively against measure logic, but are otherwise clinically relevant and realistic, e.g. scenarios where an HIV diagnosis was not documented or an HIV viral load was performed without a documented result. Negative testing further validates measure logic by accurately evaluating patients against expected outcomes and simulating the effect of missing data on measure results.

In addition to Bonnie testing, the measure specifications were reviewed independently by three eCQM experts to confirm the logic was syntactically correct, using appropriate and current versions of the eCQM standards and terminologies, and consistent with the intent of the chart-abstracted measure.

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

Bonnie testing results provide logic coverage and passing rates. The synthetic bundle reached 100% coverage, confirming each logic pathway was tested. The results also showed 100% passing rate, confirming all synthetic patients performed as expected.

Full details on Bonnie testing results are contained in the Bonnie testing attachment. The attachment includes a human-readable (HTML) summary document that lists each patient within the bundle and its passing status against expected measure outcomes. The attachment also includes a summary spreadsheet for the synthetic patient bundle which lists each patient, associated demographics, expected and actual measure population outcomes, and which portions or each measure population logic the patient meets expectations for.

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?*)

The results of measure logic testing through use of Bonnie provided confidence in the measure logic accurately representing the clinical intent and alignment with the chart abstracted measure.

2b3. EXCLUSIONS ANALYSIS (FOR MEASURS WITH EXCLUSIONS --- gap in visits and medical visit frequency)
NA ☒ no exclusions — *skip to section ___*

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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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 ___.

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

2b4.1.1 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.

Not applicable.

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.

The Ryan White HIV/AIDS Program provides a comprehensive system of care that includes primary medical care and essential support services for people living with HIV who are uninsured or underinsured. The Program works with cities, states, and local community-based organizations to provide HIV care and treatment services to more than half a million people each year. The Program reaches approximately 52% of all people diagnosed with HIV in the United States.

As indicated in data presented earlier, the Ryan White HIV/AIDS Program is a public health, safety net program providing care to a high proportion of racial/ethnic minority, transgender, unstable housing, and low income people living with HIV. Many of people served by the Ryan White HIV/AIDS Program represent sociodemographics factors incorporate in risk adjusting models by many measures stewards. As a result, the Ryan White HIV/AIDS Program does not adjust for risk in its performance measures. Rather, it is a fundamental aspect of the Ryan White HIV/AIDS Program to identify disparities and work to improve quality of care for subpopulations. Additionally, this measure is not used for pay-for-performance, bonuses, or penalties.

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*)

Not applicable.

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

Not applicable.

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)

Not applicable.

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

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

Not applicable.

If stratified, skip to ____

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:** Not applicable.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*): Not applicable.

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves: Not applicable.

2b4.9. Results of Risk Stratification Analysis: Not applicable.

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*) Not applicable.

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*)

Not applicable.

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)*

The chart-abstracted version of this measure has been in use since 2010. To examine meaningful differences in performance, we examined the distribution of the proportion of patients with viral suppression across providers, by year. Performance scores were broken into the bottom 10% and top 90% providers to better characterize the gaps that remain across providers. Moreover, performance scores were examined with respect to NHAS 2020 Indicator 6: increase the percentage of persons with diagnosed HIV infection who are virally suppressed to at least 80 percent.

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*)

	% Patients with viral suppression across providers					Providers achieving ≥80% suppression		
Year	Mean	SD	Median	10th %ile	90th %ile	N	n	%
2010	60.6	23.8	67.8	19.5	82.8	846	145	17.1
2011	64.7	22.1	71.4	31.9	84.9	811	207	24.5
2012	69.9	20.3	75.6	40.2	88.0	816	277	32.7
2013	76.1	17	80.7	57.1	90.2	823	435	51.4
2014	80.3	15.5	84.2	65.0	93.1	813	530	65.2

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. In 2014, the bottom 10% of providers had viral suppression rates of 65.0% or lower; the top 90% of providers had viral suppression rates of 93.1% or higher. While this gap appears to be

narrowing over time, a meaningful difference of 28.1 percentage points remains, demonstrating the value of the measure in identifying sites based on poor performance relative to the top performers.

Provider-level performance differences observed in the table above also underscore improvements in the proportion of patients with viral suppression in achieving 80% viral suppression. In 2014, of 813 providers, 530 (65.2%) had at least 80% of patients reach viral suppression. Additionally, the overall percentage of patients with viral suppression was 80.3%; however, given the large population that the RWHAP serves, even the poorest performing sites (e.g., bottom 10%) represent a substantial number of patients.

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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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*)

The HQMF standard specifies that if data are unknown or missing, they shall fail the criterion. This constraint embodies the notion that absence of evidence is evidence of absence, i.e. data not present in a structured field from which the measure draws will not be considered for measure calculation. In certain cases, missing data may have no impact on the measure outcome for a given patient. For example, a data element used in a series of OR statements will not impact the measure outcome if another data element in the OR statement is present and meets all other defined constraints.

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*)

The Bonnie synthetic patient bundle includes scenarios for missing data elements, which are a form of negative testing. All Bonnie synthetic patients with missing data performed according to the HQMF standard specification and as expected.

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*)

Please see response for question 2b7.1 above.

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), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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: NQFXXX_HIVViralSuppression_Feasibility_Scorecard_v1.0-636177547747228995.xlsx

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.
Not applicable.

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)*.
The measure specifications contain limited proprietary codes for convenience. Users of CPT(R) should obtain all necessary licenses from the owners of these code sets.
The use of SNOMED Clinical Terms(R) requires a Unified Medical Language System (UMLS) license. These licenses are freely available, from the National Library of Medicine.

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 Payment Program	

Quality Improvement (external benchmarking to organizations) Quality Improvement (Internal to the specific organization)	

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

Medicaid Adult Core Set
Sponsor: Federal government
Geographic area: Nationwide
Accountable entities: State Medicaid programs
Patients: Unknown

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

National HIV/AIDS Strategy
Sponsor: Federal government
Geographic area: Nationwide
Accountable entities: Federal agencies and service providers
Patients: All people living with HIV in the United States

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.*)
This measure is current under consideration for the Centers for Medicare and Medicaid Merit Based Incentive Payment System (MIPS).

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.
HIV viral suppression has been improving in the United States since the first release of publically available data. The Ryan White HIV/AIDS Program served more than 300,000 unduplicated patients annually between 2010-2014 across 2,000+ grant recipients and subrecipients. The Ryan White HIV/AIDS Program has experienced a 20-point increase in viral suppression from 61.8% in 2010 to 80.3% in 2014. Viral suppression has increased across all demographic groups and subpopulations.

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.**
N/A

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.
Starting in 2015, Health Resources and Services Administration began releasing December 1st – World AIDS Day – an annual data report (Ryan White HIV/AIDS Program Annual Client-Level Data Report) that contains data similar to those presenting in the report. Building upon the success of the state profiles (http://hab.hrsa.gov/stateprofiles/), Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived). A supplemental report exploring data for the eligible metropolitan areas and transitional grant areas and youth/young adults has been released as well as slides sets for fact sheets by program and population, special populations (http://hab.hrsa.gov/publications/hivaids-bureau-fact-sheets), and infographics (contained in fact sheets). Additionally, grant recipient level reports are prepared and disseminated to all Ryan White HIV/AIDS Program grant recipients.

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.**
Starting in 2015, Health Resources and Services Administration began releasing December 1st – World AIDS Day – an annual data report (Ryan White HIV/AIDS Program Annual Client-Level Data Report) that contains data similar to those presenting in the report. Building upon the success of the state profiles (http://hab.hrsa.gov/stateprofiles/), Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived). A supplemental report exploring data for the eligible metropolitan areas and transitional grant areas and youth/young adults has been released as well as slides sets for fact sheets by program and population, special populations

(http://hab.hrsa.gov/publications/hivaids-bureau-fact-sheets), and infographics (contained in fact sheets). Additionally, grant recipient level reports are prepared and disseminated to all Ryan White HIV/AIDS Program grant recipients.

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 from Ryan White HIV/AIDS Program grant recipients and subrecipients regarding the feasibility and usefulness of the data presented in the Ryan White HIV/AIDS Program Annual Client-Level Data Report. Significant feedback has been provided about the timeliness and expansions of the data release. Grant recipient report using the data for benchmarking their program, setting goals/targets, and gaining a fuller understanding of all aspects of the Ryan White HIV/AIDS Program (i.e. other regions of the country). Grant recipients and subrecipients have also requested additional analyses. Health Resources and Services Administration responded with supplemental reports (Ryan White HIV/AIDS Program Supplemental Client-Level Data Report, Eligible Metropolitan Areas and Transitional Grant Areas; special population reports); slide decks for the overall client population and special populations; grant recipient reports; and infographics – all of which will be updated and released annually. Health Resources and Services Administration plans to release additional analyses and special reports this year based on feedback from Ryan White HIV/AIDS Program grant recipients and subrecipients.

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

4d2.3. **Summarize the feedback obtained from other users**
Ryan White HIV/AIDS Program national partners (national organizations that represent grant recipients, subrecipients, and patients) has provided antidotal feedback regarding the timeliness, feasibility, and usability of the release of the Ryan White HIV/AIDS Program Annual Client-Level Data Report, supplemental reports, slide decks, fact sheets, and infographics. The national partners encourage the continued release of the data in all its formats.

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.

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)**

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
2079 HIV Medical Visit Frequency
2080 Gap in HIV Medical Visits

2082 HIV Viral Suppression
2083 Prescription of HIV Antiretroviral Therapy
3211 Prescription of HIV Antiretroviral Therapy
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?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Harmonization exists with all measures except 405 and 409. Plan 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.)
None

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.

No appendix **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

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 #: 2083
Measure Title: Prescription of HIV Antiretroviral Therapy
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 prescribed antiretroviral therapy for the treatment of HIV infection during the measurement year. 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: Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. HIV antiretroviral therapy delays this progression and increases the length of survival.

Current HIV treatment guidelines now recommend universal prescription of HIV antiretroviral therapy for sustained viral load suppression which in turn is directly related to reduction in disease progression and reduction in potential for transmission of HIV infection. Among persons in care, sustained viral load suppression represents the cumulative effect of prescribed therapy, ongoing monitoring, and patient adherence. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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 from the denominator prescribed HIV antiretroviral therapy during the measurement year.
Denominator Statement: Number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the measurement year
Denominator Exclusions: There are no patient exclusions.

Measure Type: Process
Data Source: Process
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 [year]**

- Evidence and clinical guidelines state that Antiretroviral Therapy is recommended for all HIV-infected individuals in order to reduce morbidity and mortality. Evidence focuses on the percent of providers prescribing ART and the percent of patients with viral load suppression across those providers, the data suggests a positive correlation.
- As a whole, the general evidence suggests that prescription to ART for those infected with HIV will lead to viral suppression if treatment is maintained.

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 rationale provided for this measure is that HIV retroviral therapy (ART) delays the progression of the disease and increases the length of survival for the patient.
- The most recent data from 2014 estimates that although 86% of people living with HIV have been diagnosed, only 40% are engaged in care and only 37% have been prescribed HIV antiretroviral therapy.

Exception to evidence

N/A

Questions for the Committee:

If the developer provided updated evidence for this measure:

- *The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?*
  - *What is the relationship of this measure to patient outcomes?*
  - *How strong is the evidence for this relationship?*

- *Is the evidence directly applicable to the process of care being measured?*

- *For possible exception to the evidence criterion:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable for prescription of HIV antiretroviral therapy without empirical evidence?*

Guidance from the Evidence Algorithm

Process measure evidence based (Box3) → Empirical Evidence is unrelated to distal process of ART prescription (BOX 7)→ Possible related process measures (Box 10)→ No exception → Rate as Insufficient

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

RATIONALE: Evidence provided by the developer lacked a systematic review of the evidence and was not directly related to the process measure 2083 but to other steps in the model that lead to viral suppression in patients with HIV.

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.

Provider-level performance scores for antiretroviral treatment (ART) for 2014 are presented below.

	2014	2013	2012	2011	2010
Rate	77.6	77.5	74.3	71.1	68.4
Pts w/ ≥1 medical visit (den)	316,087	327,618	335,408	327,744	324,455
Pts prescribed ART (Num)	245,400 (77.6)	253,972 (77.5)	249,094 (74.3)	233,132 (71.1)	221,908 (68.4)
Mean	78.0	77.5	73.4	70.1	65.9
Median	90.0	86.5	83.8	79.8	76.5
Standard Deviation	28.0	24.1	25.4	26.4	27.5
10th percentile	29.6	42.9	31.7	26.1	17.8
90th percentile	98.3	96.4	94.7	93.2	91.2
Min, Max	0.0, 100.0	0.0, 100.0	0.0, 100.0	0.0, 100.0	0.0, 100.0
# of facilities	813	823	816	811	846

Disparities

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. Descriptive characteristics are provided by the developer in the table below. The full table can be found here.

	2014 N (%)	2013 N (%)	2012 N (%)	2011 N (%)	2010 N (%)
Race/Ethnicity					
Am. Indian/Alaska Native	1,272 (0.4)	1,414 (0.5)	1,371 (0.4)	1,366 (0.4)	1,473 (0.5)
Asian	3,791 (1.2)	3,835 (1.2)	3,980 (1.2)	3,598 (1.2)	3,382 (1.1)
Black/African Am.	142,746 (46.9)	146,056 (47.0)	150,974 (47.2)	149,834 (47.8)	146,460 (47.3)
Hispanic/Latino	74,714 (24.5)	74,967 (24.1)	75,201 (23.5)	71,240 (22.7)	71,002 (22.9)
Native Hawaiian/Pacific Is.	442 (0.2)	510 (0.2)	575 (0.2)	710 (0.2)	627 (0.2)
White	75,931 (24.9)	78,953 (25.4)	83,820 (26.2)	83,061 (26.5)	83,854 (27.1)
Multiple Races	5,651 (1.9)	4,899 (1.6)	4,238 (1.3)	3,716 (1.2)	3,177 (1.0)
Gender					
Male	216,965 (70.7)	221,930 (70.7)	230,075 (70.8)	223,379 (69.9)	219,625 (69.7)
Female	87,071 (28.4)	89,212 (28.4)	92,186 (28.4)	93,687 (29.3)	93,266 (29.6)
Transgender	2,974 (1.0)	2,779 (0.9)	2,848 (0.9)	2,585 (0.8)	2,313 (0.7)

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

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

Committee pre-evaluation comments

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

1a. Evidence

*The evidence provided indicates that Antiretroviral Therapy (ART) is recommended for HIV infected (HIV+) individuals to reduce morbidity and mortality. The main "evidence" provided is Clinical Guidelines, which clearly indicate that ART is recommended for all HIV+ patients. While the evidence that the use of ART therapy improves outcomes for HIV+ individuals is implied, there is no evidence provided that patients who are prescribed (and presumably take) ART have viral load suppression and improved outcomes. Based on this fact, there is INSUFFICIENT evidence.

*There is no need for repeat discussion and vote on Evidence

1b. Performance Gap

***"**Gap: There is a clear need to improve the numbers of HIV+ patients who are prescribed ART. The currently most available data from the data provided about the population studied, shows that in 2014, 77.6% receive ART, leaving 22.4% who do not receive ART. This is much higher than stated by CDC in the 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.

Disparities: There was evidence that the population characteristics represented a diverse group of people, including age groups, race/ethnicity, and gender (including gender orientation), however I do not see any data about the performance rates by these variables. As a result I cannot say that there is a disparity gap, unless I am misunderstanding the data which were presented.

There is in general a moderate performance gap."

*there is a gap in care that warrants a national performance measure

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):** • Abstracted from paper records and electronic health records **Specifications:** • This measure is specified at hospital/facility/agency level • Patients are included in the numerator if they were prescribed HIV antiretroviral therapy during the measurement year • The denominator includes the number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the measurement year • There are no patient exclusions • 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:

- Each year from 2010-2014 more than 91% of providers had reliability scores of 0.9 or greater. Therefore, the reliability of viral suppression can be considered to be sufficient to identify real differences in performance across providers. Median reliability was consistently 0.99 during 2010-2014, supporting the conclusion that the reliability of this measure can be considered very good.

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 developer used the a beta binomial model to estimate reliability, this method was calculated using the NCQA technical report "The reliability of Provider Profiling: A tutorial". The beta binomial model is appropriate for pass/fail measures according to the developer. 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.

Results of reliability testing

- Median reliability was consistently 0.99 during 2010-2014, supporting the conclusion that the reliability of this measure can be considered very good.

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?*
- *Is the 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 testing (Box 2) → Testing of the measure score (Box 4) → Appropriate method (Box 5) → High certainty (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:

- *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:

- Face validity for the measure was established through a technical work group empaneled for the development of the measure.

Describe any updates to validity testing:

- **N/A**

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:

- 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 developer assessed Face validity through a technical work group empaneled for the development of the measure. The work group voted on importance, ability to assess quality care, feasibility to implement measure, 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.
- 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.

Validity testing results:

- The technical work group agreed upon a measure that could assess and improvement the quality of HIV care.
- No comments were received that the measure should be discontinued.

Questions for the Committee:

2b3-2b7. Threats to Validity
2b3. Exclusions: No exclusions
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)*: • The Data represents variability across providers, In 2014, the bottom 10% of providers had ART prescription rates of 29.6% or lower; the top 90% of providers had rates of 98.3% or higher. These differences demonstrate

the continued value of the measure in identifying sites based on poor performance relative to the top performers.

% patients with viral suppression across provider						Providers with ≥80% patients prescribed ART		
Year	Mean	SD	Median	10th percentile	90th percentile	N	n	%
2014	78%	28%	90%	29.6%	98.3%	813	565	69.5
2013	77.5%	24.1%	86.5%	42.9%	96.4%	823	532	64.6
2012	73.4%	25.4%	83.8%	31.7%	94.7%	816	471	57.7
2011	70.1%	26.4%	79.8%	26.1%	93.2%	811	402	49.6
2010	65.9%	27.5%	76.5%	17.8%	91.2%	846	353	41.7

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- Based on the method used to calculate the ART performance score, conducting missing data analysis is not applicable for this measure.

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

*"No concerns. I believe there is a HIGH level of reliability.

All date elements well defined. The evidence presented is adequate in my minds eye, for these purposes. Test sample is adequate. There is sufficient reliability based on the information provided."

*All the data elements are clearly defined

the logic or calculation algorithm is clear
It is likely this measure can be consistently implemented"

2a2. Reliability Testing
*N/A

*there is no need to re-discuss and re-vote on reliability
the test sample is adequate to generalize for widespread implementation
the results demonstrate sufficient reliability so that differences in performance can be identified

2b1. Validity Specifications
*No Empirical validity testing was performed on the measure score, so this is rated as INSUFFICIENT. Only Face Validity was performed.

*It is necessary that antiretroviral therapy be prescribed, but it is not clear that patients actually receive that therapy from this measure.

2b2. Validity Testing
*N/A

*There is no need to re-discuss [and re-vote] on validity [testing for validity].

2b3-7 Threats to Validity
*N/A

*2b.5 the measure identifies meaningful differences about quality
"

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 the required data elements are available in electronic health records or other electronic sources and are in defined fields.
- The operational use of this measure are readily available within patient health records and provided annually to the Ryan White HIV/AIDS Program. Because of availability, sampling is not performed.

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?*
- *Is the data collection strategy ready to be put into operational use?*

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

Committee pre-evaluation comments
Criteria 3: Feasibility

3. Feasibility
*All elements are available. All available in electronic form, as of 2014. No concerns. HIGH Feasibility.

*The required data elements are routinely generated and used during care delivery
the required data elements are available in electronic form, e.g., EHR or other electronic sources?
The data collection strategy is ready to be put into operational use"

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

- 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 (PQRS) and Value Based Modifier
  - Sponsor: Federal government
  - Geographic area: Nationwide
  - Accountable entities: Physicians and practitioners
  - Patients: Unknown
- National HIV/AIDS Strategy
  - Sponsor: Federal government
  - Geographic area: Nationwide
  - Accountable entities: Federal agencies and service providers
  - Patients: All people living with HIV in the United States

Improvement results

- The developer reports that the percent of patients being prescribed ART from 2010 to 2014 has increased from 68.4 to 77.6 percent. The Ryan White HIV/AIDS Program has experienced a 10 + point increase in viral suppression from 65.9% in 2010 to 78.0% in 2014. Prescription of HIV antiretroviral therapy has increased across all demographic groups and subpopulations.

Unexpected findings (positive or negative) during implementation

- 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, prescription of HIV antiretroviral therapy is one of five stages of the HIV care continuum. This measure has become the standard when measuring prescription of HIV antiretroviral therapy.

Potential harms

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

Vetting of the measure

- Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived).
- Ryan White HIV/AIDS Program national partners (national organizations that represent grant recipients, subrecipients, and patients) has provided antidotal feedback regarding the timeliness, feasibility, and usability of the release of the Ryan White HIV/AIDS Program Annual Client-Level Data Report, supplemental reports, slide decks, fact sheets, and infographics. The national partners encourage the continued release of the data in all its formats.
- During the initial development of the chart-abstracted 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 measures are reviewed for clinical relevance, change in scientific acceptability, and consistency with guidelines. The chart-abstracted measure has not been modified as a result of the annual reviews.

Feedback:

- Anecdotal feedback has been received from Ryan White HIV/AIDS Program grant recipients and subrecipients regarding the feasibility and usefulness of the data presented in the Ryan White HIV/AIDS Program Annual Client-Level Data Report. The national partners encourage the continued release of the data in all its formats.

- Significant feedback has been provided about the timeliness and expansions of the data release. Grant recipient report using the data for benchmarking their program, setting goals/targets, and gaining a fuller understanding of all aspects of the Ryan White HIV/AIDS Program (i.e. other regions of the country). Grant recipients and subrecipients have also requested additional analyses. Health Resources and Services Administration responded with supplemental reports (Ryan White HIV/AIDS Program Supplemental Client-Level Data Report, Eligible Metropolitan Areas and Transitional Grant Areas; special population reports); slide decks for the overall client population and special populations; grant recipient reports; and infographics – all of which will be updated and released annually. Health Resources and Services Administration plans to release additional analyses and special reports this year based on feedback from Ryan White HIV/AIDS Program grant recipients and subrecipients.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *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

*Vetter in "real world" settings; public reporting of data; feedback solicited; Feedback was considered though. Feedback has been anecdotal. - MODERATE Usability and Use.

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 – population but different measurement periods
  - 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
  - 2079 HIV Medical Visit Frequency

Harmonization

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

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 candidate 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**

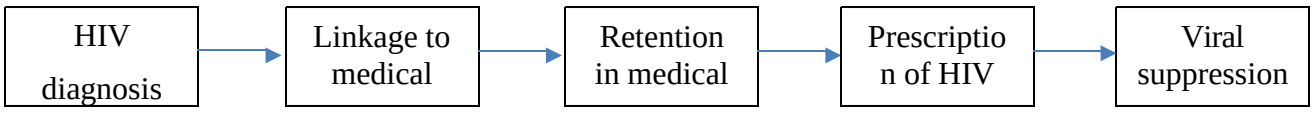
RATIONALE IF NOT ELIGIBLE: The measure is not eligible for Endorsement+ because empirical reliability and validity testing of the measure score was not conducted and the measure has not been vetted in real world settings by those being measured and other users.

Pre-meeting public and member comments

-

Measure Title: Prescription of HIV Antiretroviral Therapy

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.

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)

Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. HIV antiretroviral therapy delays this progression and increases the length of survival.

Current HIV treatment guidelines now recommend universal prescription of HIV antiretroviral therapy for sustained viral load suppression which in turn is directly related to reduction in disease progression and reduction in potential for transmission of HIV infection. Among persons in care, sustained viral load suppression represents the cumulative effect of prescribed therapy, ongoing monitoring, and patient adherence. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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.

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 15, 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 15, 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 15, 2016. http://www.iapac.org/uploads/JIAPAC-IAPAC-Guidelines-for-Optimizing-the-HIV-Care-Continuum-Supplement-Nov-Dec-2015.pdf

Günthard HF, Saag MS, Benson CA, del Rio C, Eron JJ, Gallant JE, Hoy JF, Mugavero MJ, Sax PE, Thompson MA, Gandhi RT, Landovitz RJ, Smith DM, Jacobsen DM, Volberding PA. Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults 2016 Recommendations of the International Antiviral Society–USA Panel. JAMA. 2016. https://www.iasusa.org/content/antiretroviral-drugs-treatment-and-prevention-hiv-infection-adults-2016-recommendations

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

Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents:

Initiation of Antiretroviral Therapy (page E-1)

- Antiretroviral therapy (ART) is recommended for all HIV-infected individuals, regardless of CD4 T lymphocyte cell count, to reduce the morbidity and mortality associated with HIV infection (AI).
- ART is also recommended for HIV-infected individuals to prevent HIV transmission (AI).
- When initiating ART, it is important to educate patients regarding the benefits and considerations regarding ART, and to address strategies to optimize adherence. On a case-by-case basis, ART may be deferred because of clinical and/or psychosocial factors, but therapy should be initiated as soon as possible.

Considerations for Antiretroviral Use in Special Patient Populations: Acute and Recent (Early) HIV Infection (page I-1)

- Antiretroviral therapy (ART) is recommended for all individuals with HIV-1 infection (AI) including those with early HIV-1 infection.

HIV-Infected Adolescents and Young Adults (page I-8):

- ART is recommended for all HIV-infected individuals (AI) to reduce morbidity and mortality. Thus, ART is also recommended for ART-naive adolescents. However, before initiation of therapy, adolescents' readiness and ability to adhere to therapy within their psychosocial context need to be carefully considered as partner of therapeutic decision making (AIII).

HIV-Infected Women (page I-20):

- Antiretroviral therapy (ART) is recommended for all HIV-infected women to improve their health and to reduce the risk of HIV transmission to HIV-uninfected sex partners (AI).

HIV/Hepatitis C Virus Coinfection (page J-6):

- Antiretroviral therapy (ART) may slow the progression of liver disease by preserving or restoring immune function and reducing HIV related immune activation and inflammation. For most HCV/HIV-coinfected patients, including

those with cirrhosis, the benefits of ART outweigh concerns regarding drug-induced liver injury. Therefore, ART should be initiated in all HCV/HIV-coinfected patients, regardless of CD4 T lymphocyte (CD4) cell count (AI).

WHO:

4.3 When to start ART (page xxxi)

4.3.1 When to start ART in adults (>19 years old)

- ART should be initiated in all adults living with HIV, regardless of WHO clinical stage and at any CD4 cell count (strong recommendation, moderate-quality evidence).
- As a priority, ART should be initiated in all adults with severe or advanced HIV clinical disease (WHO clinical stage 3 or 4) and adults with a CD4 count ≤350 cells/mm3 (strong recommendation, moderate-quality evidence).

4.3.2 When to start ART in pregnant and breastfeeding women

- ART should be initiated in all pregnant and breastfeeding women living with HIV, regardless of WHO clinical stage and at any CD4 cell count and continued lifelong (strong recommendation, moderate-quality evidence).

4.3.3 When to start HIV antiretroviral therapy in adolescents (10–19 years of age)

- ART should be initiated in all adolescents living with HIV, regardless of WHO clinical stage and at any CD4 cell count (conditional recommendation, low-quality evidence).
- As a priority, ART should be initiated in all adolescents with severe or advanced HIV clinical disease (WHO clinical stage 3 or 4) and adolescents with a CD4 count ≤350 cells/mm3 (strong recommendation, moderate-quality evidence).

4.3.4 When to start HIV antiretroviral therapy in children younger than 10 years of age

- ART should be initiated in all children living with HIV, regardless of WHO clinical stage or at any CD4 cell count:
- Infants diagnosed in the first year of life (strong recommendation, moderate-quality evidence).
- Children living with HIV 1-year-old to less than 10 years old (conditional recommendation, low-quality evidence).
- As a priority, ART should be initiated in all children <2 years of age or children younger than 5 years of age with WHO clinical stage 3 or 4 or CD4 count ≤750 cells/mm³ or CD4 percentage <25% and children 5 years of age and older with WHO clinical stage 3 or 4 or CD4 count ≤350 cells/mm³ (strong recommendation, moderate-quality evidence).

4.3.5 Timing of HIV ANTIRETROVIRAL THERAPY for adults and children with TB

- ART should be started in all TB patients living with HIV regardless of CD4 count (strong recommendation, high-quality evidence).

International Advisory Panel on HIV Care Continuum Optimization (IAPAC):

Increasing HIV treatment coverage (page 3)

- The immediate offer of ART after HIV diagnosis, irrespective of CD4 count or clinical stage, is recommended. (AI)

Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults 2016 Recommendations of the International Antiviral Society–USA Panel

Box 1. Recommendations for When to Start (page 193)

- Antiretroviral therapy (HIV ANTIRETROVIRAL THERAPY) is recommended for all viremic patients with established HIV infection, regardless of CD4 cell count (evidence rating AIa).
- Initiation of ART is recommended as soon as possible in the setting of acute HIV infection (evidence rating BIII).
- Planned discontinuation of early ART after a specific duration of treatment is not recommended outside a research setting (evidence rating AIa).
- Initiation of ART is recommended for individuals who have persistent undetectable viral load without ART but have declining CD4 cell counts (evidence rating BIII).

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:

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

Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults, 2016 Recommendations of the International Antiviral Society–USA Panel:

Table 1. Strength of Recommendation and Quality of Evidence Rating Scale

Rating	Definition
Strength of recommendation	
A	Strong support for the recommendation
B	Moderate support for the recommendation
C	Limited support for the recommendation
Quality of evidence	
Ia	Evidence for ≥ 1 randomized clinical trials published in the peer-reviewed literature
Ib	Evidence for ≥ 1 randomized clinical trials presented in abstract form at peer-reviewed scientific meetings
IIa	Evidence from nonrandomized clinical trials or cohorts or case-control studies published in the peer-reviewed literature
IIb	Evidence from nonrandomized clinical trials or cohorts or case-control studies published in the peer-reviewed scientific meeting
III	Recommendation based on panel's analysis of the accumulated available evidnce

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

ART_evidence_NQF-636174955634964398.docx,ART_submission_form-636179052221226279.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.

Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. HIV antiretroviral therapy delays this progression and increases the length of survival.

Current HIV treatment guidelines now recommend universal prescription of HIV antiretroviral therapy for sustained viral load suppression which in turn is directly related to reduction in disease progression and reduction in potential for transmission of HIV infection. Among persons in care, sustained viral load suppression represents the cumulative effect of prescribed therapy, ongoing monitoring, and patient adherence. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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 "ART 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 "ART 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:** ART_Data_dictionary-636179051636713033.docx

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 from the denominator prescribed HIV antiretroviral therapy during the measurement year.

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 were prescribed HIV antiretroviral therapy during the measurement year. HIV antiretroviral therapy at least one HIV antiretroviral medication.

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 measurement year

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 year
2. Patients diagnosed with HIV during the first 3 months of the measurement year or prior to the measurement year
3. Patients who had at least one medical visit during the measurement year

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
There are no patient exclusions.

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.)*
There are no patient exclusions.

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.)*
N/A

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 measurement year or prior to the measurement year; and 2.) had at least one medical visit during the measurement year. The individuals who met these criteria are the denominator population.
2. Identify the individuals from the denominator population who meet the criterion for inclusion in the numerator: prescribed HIV antiretroviral therapy during the measurement year.
3. Calculate the percentage 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.
Other, Paper Records, Pharmacy

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.
Not applicable.

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
ART_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.)
Yes

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*)**:** 2083
Measure Title: Prescription of HIV Antiretroviral Therapy
Date of Submission: Click here to enter a date
Type of Measure: Process

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

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: **(*must be consistent with data sources entered in S.23*)**	**Measure Tested with Data From:**
☒ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

On an annual basis, Ryan White HIV/AIDS Program (RWHAP) grant recipient and subrecipients submit the Ryan White HIV/AIDS Services Report (RSR). The RSR dataset is the Health Resources and Services Administration HIV/AIDS Bureau's primary source of annual, client-level data collected from its nearly 2,000 funded grant recipients and subrecipients. Since 2010, client-level RSR data have been used to assess the numbers and types of clients receiving services and their HIV outcomes. Project Officers at the HIV/AIDS Bureau share the data with grant recipients and subrecipients to monitor and support their progress at improving care and treatment for people living with HIV. It is through the hard work of these providers and the RWHAP community that clients are helped every day.

RSR includes all clients served by the RWHAP during calendar years 2010 through 2014. RSR data do not include information about AIDS Drug Assistance Programs (ADAP); all ADAP-related information is collected through another data system. Although data presented in this report are "nonADAP," this does not imply the clients did not receive ADAP services. ADAP data will be published separately, at later time.

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 <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

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

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

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. Over 800 (varies by year) Ryan White HIV/AIDS Program outpatient ambulatory medical care providers representing various types, locations, and sizes were included in the testing.

Descriptive characteristics of RWHAP providers

	2010		2011		2012		2013		2014	
	N	%	N	%	N	%	N	%	N	%
Overall	846		811		816		823	--	813	--
Provider type										
Hospital or university-based clinic	355	17.5	358	18.6	349	19.1	351	19.6	338	19.4
Community based organization	1,114	54.9	1,053	54.8	993	54.3	958	53.6	921	53.0
Health department	284	14.0	274	14.3	243	13.3	233	13.0	243	14.0
Other	275	13.6	237	12.3	243	13.3	247	13.8	237	13.6
HHS Region										
Region 1	149	8.0	153	8.6	142	8.4	139	8.4	135	8.3
Region 2	368	19.7	339	19.0	323	19.1	303	18.3	293	18.1
Region 3	180	9.6	177	9.9	174	10.3	174	10.5	160	9.9
Region 4	337	18.0	335	18.8	312	18.5	301	18.1	313	19.3
Region 5	197	10.5	189	10.6	177	10.5	188	11.3	180	11.1
Region 6	150	8.0	142	8.0	133	7.9	131	7.9	132	8.2
Region 7	65	3.5	60	3.4	57	3.4	56	3.4	54	3.3
Region 8	48	2.6	43	2.4	34	2.0	35	2.1	46	2.8
Region 9	300	16.0	281	15.7	277	16.4	276	16.6	253	15.6
Region 10	78	4.2	68	3.8	60	3.6	56	3.4	52	3.2

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.

Distribution of patients per provider by year, 2010-2014

Year	N patients, mean	N patients, median	Min patients	Max patients
2010	384	177	1	13,159
2011	404	182	1	13,380
2012	411	179	1	13,849
2013	398	181	1	14,755
2014	388	177	1	13,850

Descriptive characteristics of RWHAP patients by year, 2010-2014

	2010		2011		2012		2013		2014	
	No.	%	No.	%	No.	%	No.	%	No.	%
OVERALL	324,455	–	327,744	–	335,408	–	327,618	–	316,087	–
AGE GROUP										
<13	3,709	1.2	3,647	1.1	3,150	1.0	2,667	0.9	2,720	0.9
13–14	627	0.2	605	0.2	469	0.1	360	0.1	343	0.1
15–19	3,698	1.2	3,541	1.1	3,066	0.9	2,609	0.8	2,506	0.8
20–24	14,040	4.5	14,831	4.6	15,741	4.8	15,538	5.0	14,578	4.8
25–29	22,120	7.0	23,278	7.3	24,904	7.7	25,586	8.2	26,043	8.5
30–34	28,644	9.1	29,330	9.2	30,084	9.3	29,495	9.4	28,484	9.3
35–39	35,161	11.2	33,597	10.5	33,005	10.2	31,560	10.1	30,691	10.0
40–44	50,769	16.1	47,941	15.0	45,343	14.0	40,728	13.0	37,000	12.1
45–49	60,344	19.2	59,453	18.6	58,145	17.9	52,863	16.8	47,932	15.6
50–54	46,433	14.7	48,647	15.2	50,876	15.7	50,491	16.1	50,492	16.4
55–59	28,015	8.9	30,646	9.6	33,215	10.2	33,493	10.7	34,667	11.3
60–64	13,441	4.3	15,237	4.8	16,991	5.2	17,780	5.7	19,399	6.3
≥65	8,187	2.6	8,946	2.8	10,147	3.1	10,780	3.4	12,231	4.0
RACE/ETHNICITY										
American Indian/ Alaska Native	1,473	0.5	1,366	0.4	1,371	0.4	1,414	0.5	1,272	0.4
Asian	3,382	1.1	3,598	1.2	3,980	1.2	3,835	1.2	3,791	1.2
Black/ African American	146,460	47.3	149,834	47.8	150,974	47.2	146,056	47.0	142,746	46.9
Hispanic/Latino[a]	71,002	22.9	71,240	22.7	75,201	23.5	74,967	24.1	74,714	24.5
Native Hawaiian/ Pacific Islander	627	0.2	710	0.2	575	0.2	510	0.2	442	0.2
White	83,854	27.1	83,061	26.5	83,820	26.2	78,953	25.4	75,931	24.9
Multiple races	3,177	1.0	3,716	1.2	4,238	1.3	4,899	1.6	5,651	1.9

GENDER										
Male	219,625	69.7	223,379	69.9	230,075	70.8	221,930	70.7	216,965	70.7
Female	93,266	29.6	93,687	29.3	92,186	28.4	89,212	28.4	87,071	28.4
Transgender	2,313	0.7	2,585	0.8	2,848	0.9	2,779	0.9	2,974	1.0
TRANSMISSION RISK										
Male client										
Male-to-male sexual contact	117,267	59.9	120,622	60.2	128,744	61.8	127,571	62.2	127,624	62.7
Injection drug use	17,479	8.9	16,787	8.4	15,586	7.5	15,509	7.6	13,753	6.8
Male-to-male sexual contact and injection drug use	6,971	3.6	6,837	3.4	6,974	3.3	6,136	3.0	6,396	3.1
Heterosexual contact	48,903	25.0	50,814	25.4	52,266	25.1	51,174	24.9	51,155	25.1
Perinatal infection	3,830	2.0	3,919	2.0	3,604	1.7	3,419	1.7	3,456	1.7
Other	1,248	0.6	1,231	0.6	1,309	0.6	1,402	0.7	1,189	0.6
Female client										
Injection drug use	9,264	11.2	9,022	10.7	8,182	9.8	8,310	10.0	7,396	9.1
Heterosexual contact	68,009	82.4	69,767	82.8	70,362	84.1	69,356	83.9	69,090	84.8
Perinatal infection	4,338	5.3	4,587	5.4	4,182	5.0	4,003	4.8	4,093	5.0
Other	900	1.1	877	1.0	936	1.1	1,044	1.3	940	1.2
Transgender client										
Sexual contact	1,874	90.7	2,058	91.2	2,281	91.8	2,314	92.9	2,499	93.2
Injection drug use	38	1.8	32	1.4	35	1.4	32	1.3	31	1.2
Sexual contact and injection drug use	144	7.0	156	6.9	158	6.4	130	5.2	135	5.0
Perinatal infection	5	0.2	5	0.2	2	0.1	4	0.2	9	0.3
Other	6	0.3	5	0.2	8	0.3	10	0.4	8	0.3
HEALTH CARE COVERAGE										
Private only	35,392	12.4	37,532	12.3	39,972	12.7	37,204	12.1	–	–
Medicare only	23,245	8.1	24,279	8.0	23,538	7.5	22,840	7.5	–	–
Medicaid only	73,292	25.6	75,690	24.8	71,990	22.8	69,211	22.6	–	–
Other public	22,398	7.8	20,977	6.9	28,039	8.9	27,347	8.9	–	–
Other private	11,512	4.0	9,884	3.2	6,049	1.9	3,682	1.2	–	–
No coverage	86,220	30.1	100,001	32.8	103,150	32.7	101,524	33.1	–	–
Multiple coverages	34,276	12.0	36,330	11.9	42,969	13.6	44,578	14.6	–	–
Private employer	–	–	–	–	–	–	–	–	18,805	6.3
Private individual	–	–	–	–	–	–	–	–	16,154	5.4
Medicare	–	–	–	–	–	–	–	–	26,145	8.7
Medicaid	–	–	–	–	–	–	–	–	94,993	31.6
Medicare and Medicaid	–	–	–	–	–	–	–	–	19,207	6.4
Veterans Administration	–	–	–	–	–	–	–	–	454	0.2
Indian Health Service	–	–	–	–	–	–	–	–	71	0.0
Other plan	–	–	–	–	–	–	–	–	11,899	4.0
No coverage	–	–	–	–	–	–	–	–	90,828	30.2

Multiple coverages	–	–	–	–	–	–	–	–	22,428	7.5

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.

Ryan White HIV/AIDS Program Services Report (RSR) 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 (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

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.

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*)

Provider-level reliability results for the "prescribed ART" measure in 2014 are detailed below. Results for years 2010-2013 are available upon request, but were not included due to space constraints.

Provider-level "prescribed ART" reliability testing (signal to noise) results, 2014.

Site/provider ID	% suppressed	variance within	reliability
55	55.6%	0.001	0.992
63	91.7%	0.003	0.958
82	16.1%	0.001	0.980
88	88.7%	0.000	0.999
96	70.3%	0.000	0.995
101	45.7%	0.002	0.976
105	62.8%	0.002	0.978
112	91.8%	0.000	0.995
113	97.2%	0.000	0.998
117	96.7%	0.000	1.000
118	94.6%	0.000	0.998
120	92.8%	0.000	0.994
123	92.8%	0.000	0.998
124	95.3%	0.000	0.997
127	82.3%	0.001	0.993
128	77.9%	0.000	0.995
133	89.8%	0.000	0.999
135	92.9%	0.000	0.994
138	86.4%	0.003	0.965
140	100.0%	0.000	1.000
141	58.8%	0.001	0.985
143	76.6%	0.001	0.981
144	88.4%	0.000	0.996
147	82.7%	0.001	0.986
148	91.8%	0.002	0.979
149	89.6%	0.001	0.989
154	95.6%	0.000	0.999
155	100.0%	0.000	1.000
156	100.0%	0.000	1.000
158	70.9%	0.002	0.979
159	85.8%	0.001	0.992
160	97.8%	0.000	0.997
164	81.7%	0.002	0.972

168	97.8%	0.000	1.000
169	96.5%	0.000	0.997
170	97.3%	0.000	0.999
171	45.4%	0.001	0.988
172	97.9%	0.000	0.999
173	70.1%	0.001	0.982
174	96.8%	0.000	0.999
175	89.8%	0.001	0.992
176	95.7%	0.000	1.000
177	82.5%	0.002	0.970
178	80.3%	0.000	0.999
179	91.4%	0.001	0.987
181	94.6%	0.000	0.999
182	49.0%	0.000	0.996
183	92.7%	0.000	0.998
184	96.5%	0.000	0.999
186	100.0%	0.000	1.000
187	96.5%	0.000	0.999
188	100.0%	0.000	1.000
191	92.9%	0.000	0.998
192	30.5%	0.000	0.995
194	88.6%	0.000	0.994
196	95.0%	0.000	0.999
197	92.1%	0.000	0.999
199	87.2%	0.000	0.997
201	94.6%	0.000	1.000
203	97.5%	0.000	1.000
205	100.0%	0.000	1.000
207	95.7%	0.000	0.998
209	100.0%	0.000	1.000
210	95.8%	0.000	0.995
211	93.7%	0.000	0.998
212	92.3%	0.000	0.996
213	93.4%	0.000	0.999
214	97.1%	0.000	0.999
215	98.8%	0.000	1.000
216	92.9%	0.000	0.998
217	97.2%	0.000	0.998

220	95.0%	0.000	0.998
221	93.0%	0.000	0.995
222	7.1%	0.000	0.997
223	3.8%	0.000	1.000
224	96.8%	0.000	0.999
225	71.8%	0.000	0.999
227	95.6%	0.000	0.999
228	89.3%	0.001	0.993
230	95.9%	0.000	1.000
231	80.8%	0.000	0.997
232	85.7%	0.001	0.993
233	95.1%	0.000	0.999
235	93.1%	0.000	0.994
236	82.7%	0.000	0.994
238	97.9%	0.000	1.000
239	76.2%	0.000	0.997
240	83.3%	0.000	0.996
241	88.6%	0.000	1.000
242	92.7%	0.000	0.995
244	97.7%	0.000	0.999
245	91.3%	0.000	0.998
246	96.2%	0.000	1.000
248	84.5%	0.000	0.999
252	96.8%	0.000	0.999
253	96.2%	0.000	0.999
255	46.9%	0.003	0.966
256	89.7%	0.000	0.996
257	98.6%	0.000	1.000
259	55.0%	0.002	0.970
263	1.7%	0.000	1.000
265	97.8%	0.000	1.000
266	96.3%	0.000	1.000
267	0.6%	0.000	1.000
268	19.9%	0.001	0.993
269	96.4%	0.000	0.999
271	97.2%	0.000	0.995
273	76.4%	0.000	0.998
275	93.5%	0.000	0.999

276	90.3%	0.000	0.996
277	89.5%	0.000	0.999
278	2.2%	0.000	1.000
279	90.5%	0.000	0.999
280	93.8%	0.002	0.975
283	82.4%	0.000	0.998
284	97.2%	0.000	0.999
285	83.5%	0.001	0.992
286	91.8%	0.000	0.997
288	91.0%	0.001	0.992
289	84.9%	0.000	0.994
290	96.1%	0.000	0.993
291	96.7%	0.000	0.999
292	90.0%	0.000	0.994
294	17.1%	0.000	0.999
295	87.5%	0.000	0.995
298	86.8%	0.000	0.999
299	3.5%	0.000	0.999
302	90.6%	0.000	0.998
303	100.0%	0.000	1.000
304	95.0%	0.000	0.999
305	96.3%	0.000	1.000
307	94.7%	0.000	0.999
308	81.5%	0.001	0.981
310	93.8%	0.000	1.000
311	92.3%	0.000	0.995
312	92.5%	0.001	0.990
313	94.0%	0.000	0.995
314	96.4%	0.000	1.000
315	98.6%	0.000	1.000
316	89.2%	0.000	0.995
317	94.7%	0.000	0.997
318	83.8%	0.000	0.996
319	96.9%	0.000	0.997
320	97.5%	0.000	0.998
321	92.0%	0.001	0.993
322	91.4%	0.000	0.993
323	79.2%	0.000	0.994

324	97.3%	0.000	0.998
325	79.3%	0.000	0.999
326	98.1%	0.000	1.000
328	1.4%	0.000	1.000
329	96.8%	0.000	1.000
332	99.5%	0.000	1.000
333	93.9%	0.001	0.992
334	96.9%	0.000	0.999
335	94.1%	0.000	0.999
336	88.3%	0.000	0.997
340	94.5%	0.000	0.998
342	97.6%	0.000	0.999
343	97.4%	0.000	0.998
344	43.1%	0.000	0.995
345	95.0%	0.000	0.995
347	89.9%	0.000	0.999
348	96.6%	0.000	0.999
349	91.2%	0.000	0.997
351	96.3%	0.000	0.999
353	53.4%	0.000	0.995
357	94.2%	0.000	0.999
358	96.5%	0.000	0.997
360	93.8%	0.000	0.999
361	92.1%	0.000	0.998
362	97.1%	0.000	0.994
363	94.1%	0.000	0.994
365	88.1%	0.003	0.967
366	91.0%	0.000	0.994
368	94.8%	0.000	0.998
369	99.1%	0.000	1.000
370	86.5%	0.001	0.982
371	99.0%	0.000	0.999
372	89.7%	0.000	0.999
375	75.2%	0.000	0.999
378	92.3%	0.001	0.985
379	98.4%	0.000	1.000
380	84.3%	0.000	0.996
382	98.2%	0.000	1.000

384	86.6%	0.000	0.994
385	96.2%	0.000	0.998
386	95.5%	0.000	0.997
388	95.0%	0.000	1.000
389	81.1%	0.001	0.990
390	90.6%	0.000	0.997
391	87.3%	0.001	0.993
393	96.0%	0.000	0.999
394	76.6%	0.003	0.963
395	96.3%	0.000	0.999
400	95.7%	0.000	0.998
404	91.6%	0.001	0.993
407	17.6%	0.000	0.995
408	95.5%	0.000	0.999
409	96.0%	0.000	0.996
410	37.5%	0.003	0.961
412	91.4%	0.000	0.998
414	97.7%	0.000	0.998
417	61.6%	0.002	0.972
421	88.8%	0.000	0.998
422	96.6%	0.000	0.999
423	7.8%	0.000	0.997
425	94.0%	0.000	0.997
427	89.4%	0.000	0.999
438	93.1%	0.000	0.998
441	88.2%	0.001	0.981
457	97.0%	0.000	1.000
463	82.2%	0.000	0.994
469	83.6%	0.000	0.995
473	0.0%	0.000	1.000
480	79.6%	0.000	0.994
481	94.5%	0.000	0.998
483	83.6%	0.000	0.998
489	73.3%	0.000	0.996
491	4.5%	0.000	0.993
498	96.9%	0.000	0.999
504	95.9%	0.000	1.000
506	96.8%	0.000	0.999

509	95.2%	0.000	0.999
510	92.7%	0.001	0.993
517	81.1%	0.000	0.993
534	98.2%	0.000	1.000
553	65.2%	0.010	0.881
593	95.9%	0.000	0.996
598	88.1%	0.000	0.994
612	97.8%	0.000	0.997
664	97.2%	0.000	0.999
704	99.0%	0.000	0.999
710	84.3%	0.003	0.966
726	82.5%	0.000	0.993
738	72.2%	0.011	0.867
744	91.3%	0.000	0.997
753	75.2%	0.001	0.992
757	97.7%	0.000	0.998
762	84.2%	0.000	0.997
765	95.2%	0.000	0.998
775	79.2%	0.002	0.979
783	73.3%	0.004	0.944
787	97.4%	0.000	0.998
791	97.6%	0.000	0.999
793	90.5%	0.004	0.947
794	69.7%	0.003	0.958
798	48.1%	0.000	0.996
799	96.3%	0.001	0.991
800	94.1%	0.000	0.996
801	100.0%	0.000	1.000
803	85.7%	0.000	0.995
807	97.4%	0.000	0.999
818	99.2%	0.000	1.000
820	90.4%	0.001	0.988
821	98.8%	0.000	0.998
824	95.5%	0.001	0.987
841	98.9%	0.000	0.998
852	100.0%	0.000	1.000
861	100.0%	0.000	1.000
867	50.9%	0.000	0.996

871	76.9%	0.000	0.998
873	81.7%	0.001	0.993
894	89.2%	0.000	0.999
905	49.1%	0.000	0.996
907	72.9%	0.000	0.999
913	15.5%	0.000	0.999
920	20.0%	0.011	0.872
926	81.1%	0.004	0.946
927	2.7%	0.000	0.997
929	59.3%	0.002	0.977
933	8.8%	0.000	0.995
945	95.0%	0.000	0.995
980	85.9%	0.001	0.988
986	87.9%	0.001	0.990
992	94.4%	0.000	0.999
996	77.1%	0.000	0.997
1009	62.1%	0.002	0.979
1017	97.9%	0.000	0.997
1022	96.7%	0.000	0.998
1023	72.3%	0.001	0.986
1026	97.1%	0.000	0.996
1029	76.1%	0.001	0.985
1031	93.6%	0.000	0.998
1036	99.0%	0.000	1.000
1037	99.0%	0.000	0.999
1038	91.5%	0.000	0.997
1049	87.1%	0.000	0.996
1050	97.2%	0.000	1.000
1052	85.2%	0.005	0.940
1055	78.3%	0.001	0.993
1056	94.9%	0.000	0.999
1066	93.9%	0.000	0.999
1067	84.2%	0.000	0.997
1068	72.6%	0.000	0.998
1093	78.3%	0.002	0.967
1094	79.7%	0.000	0.997
1100	95.3%	0.000	0.999
1109	96.9%	0.000	1.000

1110	96.7%	0.000	0.996
1112	97.3%	0.000	1.000
1120	66.5%	0.000	0.996
1121	96.3%	0.000	1.000
1122	90.5%	0.000	0.997
1131	91.0%	0.000	0.999
1132	67.2%	0.001	0.984
1146	53.3%	0.017	0.814
1155	95.0%	0.000	0.998
1160	2.2%	0.000	1.000
1162	1.9%	0.000	0.995
1163	2.0%	0.000	0.999
1167	90.6%	0.000	1.000
1214	98.0%	0.000	0.998
1216	85.2%	0.001	0.984
1229	97.1%	0.000	0.999
1230	65.8%	0.002	0.979
1263	100.0%	0.000	1.000
1276	78.9%	0.001	0.993
1278	79.6%	0.002	0.979
1284	100.0%	0.000	1.000
1287	96.0%	0.000	0.999
1289	94.7%	0.001	0.982
1300	43.5%	0.004	0.948
1302	87.4%	0.000	0.995
1309	65.3%	0.003	0.959
1310	96.4%	0.001	0.992
1314	83.5%	0.000	0.997
1318	71.2%	0.003	0.963
1319	86.3%	0.001	0.983
1333	21.4%	0.001	0.992
1349	1.8%	0.000	0.996
1358	46.3%	0.001	0.981
1359	89.3%	0.000	0.999
1364	89.0%	0.001	0.982
1378	99.3%	0.000	0.999
1380	88.5%	0.001	0.982
1382	67.9%	0.002	0.979

1401	63.7%	0.000	0.995
1430	47.2%	0.005	0.939
1444	70.3%	0.006	0.928
1445	82.4%	0.000	0.999
1448	52.4%	0.012	0.860
1451	93.6%	0.001	0.991
1456	40.1%	0.000	0.998
1461	96.9%	0.000	0.997
1464	56.7%	0.001	0.988
1479	91.7%	0.006	0.920
1490	92.5%	0.002	0.977
1511	94.0%	0.000	0.999
1512	82.9%	0.001	0.989
1514	25.6%	0.004	0.943
1527	75.0%	0.009	0.886
1552	93.2%	0.000	0.999
1567	97.8%	0.000	0.997
1570	92.2%	0.000	0.994
1572	82.4%	0.003	0.962
1574	91.7%	0.001	0.989
1582	90.7%	0.001	0.993
1583	100.0%	0.000	1.000
1587	83.3%	0.012	0.863
1594	76.3%	0.003	0.960
1597	40.7%	0.009	0.891
1607	87.9%	0.000	0.994
1610	88.9%	0.000	0.994
1628	28.1%	0.006	0.920
1634	88.5%	0.001	0.987
1635	90.9%	0.001	0.990
1637	89.6%	0.000	0.996
1650	96.0%	0.000	0.996
1654	80.8%	0.003	0.961
1656	98.5%	0.000	0.998
1668	93.4%	0.000	0.999
1672	100.0%	0.000	1.000
1684	71.1%	0.000	0.998
1719	84.4%	0.000	0.999

1762	1.3%	0.000	1.000
1784	95.5%	0.000	0.996
1786	56.1%	0.004	0.951
1792	74.8%	0.000	0.997
1806	0.0%	0.000	1.000
1809	0.7%	0.000	1.000
1812	1.2%	0.000	1.000
1831	100.0%	0.000	1.000
1834	95.2%	0.001	0.985
1847	95.9%	0.000	0.995
1849	87.5%	0.001	0.990
1879	49.7%	0.001	0.982
1900	0.0%	0.000	1.000
1904	92.9%	0.002	0.968
1912	95.0%	0.000	0.999
1930	92.7%	0.001	0.992
1955	77.6%	0.002	0.980
1967	33.3%	0.007	0.915
1968	43.8%	0.008	0.904
1970	0.0%	0.000	1.000
1972	0.0%	0.000	1.000
1977	59.3%	0.009	0.891
1980	95.9%	0.001	0.989
1989	87.0%	0.001	0.986
2003	50.0%	0.063	0.538
2008	83.3%	0.023	0.759
2010	90.0%	0.009	0.890
2011	90.5%	0.004	0.947
2017	100.0%	0.000	1.000
2020	50.0%	0.025	0.744
2025	90.7%	0.000	0.996
2028	86.1%	0.001	0.992
2029	90.9%	0.001	0.987
2034	94.3%	0.001	0.986
2041	0.0%	0.000	1.000
2049	78.0%	0.004	0.946
2058	95.5%	0.000	0.998
2072	100.0%	0.000	1.000

2073	95.4%	0.000	0.999
2076	82.6%	0.006	0.921
2078	93.8%	0.002	0.975
2080	100.0%	0.000	1.000
2081	50.0%	0.125	0.368
2116	70.6%	0.002	0.975
2117	82.6%	0.006	0.921
2118	72.0%	0.002	0.971
2126	88.9%	0.001	0.990
2127	91.1%	0.001	0.988
2129	69.3%	0.002	0.980
2133	79.2%	0.000	0.994
2134	93.5%	0.002	0.974
2137	100.0%	0.000	1.000
2139	46.0%	0.002	0.979
2141	79.6%	0.000	0.994
2143	93.2%	0.000	0.999
2148	71.8%	0.002	0.978
2150	92.4%	0.000	0.999
2153	77.8%	0.001	0.985
2163	50.2%	0.001	0.984
2170	72.3%	0.001	0.991
2174	61.1%	0.000	0.999
2175	93.2%	0.000	0.995
2178	84.8%	0.000	0.996
2180	95.5%	0.000	0.999
2183	88.0%	0.001	0.992
2187	95.8%	0.002	0.978
2188	78.9%	0.009	0.893
2189	95.2%	0.002	0.971
2191	96.8%	0.001	0.993
2198	100.0%	0.000	1.000
2200	90.9%	0.003	0.967
2203	88.1%	0.001	0.987
2205	92.5%	0.001	0.982
2207	76.8%	0.000	0.999
2224	98.3%	0.000	1.000
2228	47.4%	0.013	0.847

2230	20.4%	0.001	0.981
2232	20.0%	0.032	0.695
2246	60.5%	0.001	0.986
2252	28.1%	0.003	0.958
2263	92.0%	0.003	0.961
2264	50.0%	0.125	0.368
2296	97.7%	0.000	0.999
2299	6.9%	0.000	0.998
2320	88.1%	0.000	0.999
2366	67.1%	0.003	0.958
2368	44.1%	0.004	0.953
2374	0.0%	0.000	1.000
2378	84.2%	0.007	0.912
2379	60.9%	0.010	0.875
2381	85.8%	0.001	0.989
2388	66.7%	0.037	0.663
2389	100.0%	0.000	1.000
2415	97.7%	0.000	0.998
2420	48.4%	0.002	0.973
2436	98.3%	0.000	0.999
2438	98.3%	0.000	0.996
2444	0.0%	0.000	1.000
2457	97.3%	0.000	0.998
2474	81.0%	0.001	0.992
2495	89.6%	0.001	0.981
2514	31.1%	0.000	1.000
2525	96.6%	0.000	0.999
2572	43.1%	0.000	0.995
2654	47.4%	0.007	0.917
2694	0.0%	0.000	1.000
2699	94.7%	0.003	0.965
2700	93.7%	0.000	0.998
2702	93.9%	0.000	0.998
2703	94.7%	0.000	0.999
2704	93.2%	0.000	0.997
2707	100.0%	0.000	1.000
2709	94.7%	0.000	0.999
2714	98.0%	0.000	0.999

2717	88.0%	0.001	0.989
2718	96.6%	0.000	1.000
2720	92.9%	0.001	0.993
2721	89.5%	0.000	0.999
2723	95.7%	0.000	0.998
2728	54.2%	0.001	0.987
2732	94.0%	0.000	0.998
2737	81.4%	0.004	0.954
2745	98.3%	0.000	0.998
2746	49.2%	0.002	0.973
2757	88.9%	0.000	0.994
2764	0.5%	0.000	1.000
2766	0.0%	0.000	1.000
2767	4.8%	0.001	0.985
2768	1.5%	0.000	1.000
2769	95.2%	0.001	0.990
2771	1.2%	0.000	1.000
2775	1.0%	0.000	1.000
2779	0.9%	0.000	0.999
2782	2.6%	0.000	0.998
2788	1.0%	0.000	0.999
2790	1.2%	0.000	1.000
2794	5.8%	0.001	0.986
2795	0.0%	0.000	1.000
2849	84.6%	0.000	0.995
2855	90.8%	0.001	0.991
2856	87.3%	0.002	0.973
2857	100.0%	0.000	1.000
2865	66.4%	0.000	0.995
2872	90.2%	0.002	0.971
2873	71.4%	0.029	0.714
2892	83.5%	0.001	0.993
2893	94.4%	0.000	0.997
2928	87.4%	0.000	0.997
2929	38.8%	0.003	0.963
2933	88.7%	0.001	0.992
2945	92.9%	0.000	0.996
2947	75.0%	0.000	0.997

2948	94.7%	0.000	0.999
2949	73.5%	0.001	0.984
2951	82.0%	0.001	0.990
2958	94.4%	0.001	0.980
2959	82.7%	0.003	0.964
2966	90.0%	0.001	0.990
2969	87.1%	0.002	0.978
2970	98.0%	0.000	0.997
2988	68.8%	0.002	0.969
2989	100.0%	0.000	1.000
2997	96.5%	0.000	0.999
3010	94.8%	0.000	0.998
3012	87.4%	0.001	0.992
3020	0.0%	0.000	1.000
3048	36.8%	0.012	0.856
3052	0.0%	0.000	1.000
3077	98.7%	0.000	0.998
3079	97.0%	0.000	0.997
3111	64.8%	0.003	0.958
3131	84.9%	0.001	0.992
3133	78.7%	0.002	0.975
3164	96.0%	0.001	0.990
3177	85.5%	0.001	0.990
3187	61.6%	0.003	0.957
3210	94.9%	0.000	0.998
3255	92.0%	0.001	0.987
3261	96.0%	0.000	0.997
3262	88.6%	0.002	0.970
3264	96.6%	0.001	0.984
3265	94.9%	0.000	0.996
3310	94.9%	0.000	0.997
3359	76.0%	0.001	0.980
3389	100.0%	0.000	1.000
3397	32.3%	0.001	0.987
3401	36.6%	0.002	0.972
3428	98.7%	0.000	1.000
3433	65.4%	0.001	0.985
3440	95.5%	0.000	0.998

3444	81.2%	0.001	0.990
3445	89.3%	0.002	0.977
3449	100.0%	0.000	1.000
3456	70.7%	0.001	0.992
3469	98.1%	0.000	0.999
3482	94.4%	0.000	0.993
3484	97.6%	0.000	0.999
3498	90.0%	0.005	0.942
3507	96.4%	0.001	0.991
3519	93.0%	0.000	0.997
3521	100.0%	0.000	1.000
3523	95.2%	0.000	0.995
3527	93.8%	0.000	0.997
3530	1.0%	0.000	0.999
3533	100.0%	0.000	1.000
3538	98.2%	0.000	0.998
3551	75.7%	0.001	0.992
3552	30.4%	0.009	0.888
3553	90.6%	0.002	0.978
3554	44.4%	0.009	0.888
3573	99.3%	0.000	0.999
3579	100.0%	0.000	1.000
3583	99.9%	0.000	1.000
3587	91.3%	0.001	0.986
3591	86.8%	0.003	0.960
3593	46.1%	0.001	0.986
3594	85.8%	0.001	0.987
3595	84.2%	0.000	0.998
3596	94.5%	0.000	0.994
3597	98.4%	0.000	0.999
3598	94.4%	0.001	0.990
3600	77.5%	0.004	0.943
3601	96.4%	0.000	0.999
3602	81.3%	0.002	0.978
3603	96.8%	0.000	0.997
3604	100.0%	0.000	1.000
3605	89.0%	0.001	0.991
3606	89.7%	0.000	0.993

3607	94.8%	0.000	0.999
3608	88.1%	0.000	0.994
3609	92.0%	0.001	0.980
3625	83.8%	0.002	0.973
3626	94.8%	0.001	0.989
3633	79.3%	0.002	0.973
3639	93.1%	0.000	0.999
3658	70.0%	0.021	0.776
3659	83.9%	0.001	0.980
3687	89.7%	0.001	0.991
3702	96.0%	0.000	0.999
3728	100.0%	0.000	1.000
3769	96.3%	0.000	0.999
3794	2.4%	0.000	0.999
3826	54.1%	0.002	0.979
3847	69.6%	0.000	0.997
3862	63.6%	0.021	0.776
3879	0.4%	0.000	1.000
3904	96.7%	0.000	1.000
3932	86.2%	0.001	0.982
3942	61.8%	0.002	0.976
3959	54.9%	0.003	0.954
3969	85.7%	0.002	0.971
3972	97.8%	0.000	0.998
3973	94.1%	0.000	0.997
3975	97.1%	0.001	0.989
3976	94.5%	0.000	0.998
3977	1.0%	0.000	1.000
3978	27.8%	0.000	0.995
3979	90.0%	0.000	0.998
3980	93.6%	0.000	0.997
3981	94.4%	0.000	1.000
3982	95.6%	0.000	0.999
3983	96.9%	0.000	0.999
3984	83.3%	0.003	0.966
3985	100.0%	0.000	1.000
3998	92.3%	0.001	0.991
4014	88.1%	0.001	0.989

4018	88.8%	0.001	0.987
4035	22.5%	0.001	0.983
4039	52.8%	0.002	0.969
4068	75.0%	0.009	0.886
4088	90.1%	0.000	0.995
4122	96.3%	0.000	0.994
4126	88.6%	0.001	0.981
4220	84.7%	0.001	0.987
4221	90.9%	0.001	0.985
4235	73.0%	0.002	0.974
7685	99.3%	0.000	0.999
7718	85.0%	0.002	0.979
7722	86.8%	0.002	0.980
7728	100.0%	0.000	1.000
7734	95.7%	0.000	0.998
7738	94.3%	0.000	0.999
7758	96.4%	0.001	0.983
7765	92.7%	0.000	0.998
7766	97.3%	0.000	0.999
7772	75.2%	0.000	0.994
7780	83.3%	0.023	0.759
7782	66.7%	0.012	0.855
7785	95.4%	0.000	0.998
7802	94.9%	0.000	0.999
7807	6.7%	0.000	0.994
7821	93.0%	0.001	0.987
7833	0.7%	0.000	0.999
7834	80.7%	0.001	0.986
7845	79.0%	0.001	0.990
7857	42.9%	0.035	0.675
7878	80.4%	0.003	0.955
7885	97.3%	0.000	0.999
7886	100.0%	0.000	1.000
7888	95.3%	0.000	0.998
7892	58.9%	0.002	0.978
7910	100.0%	0.000	1.000
7911	89.3%	0.003	0.955
7913	29.6%	0.003	0.961

7919	100.0%	0.000	1.000
7920	79.5%	0.002	0.970
7929	97.0%	0.000	0.999
7931	90.2%	0.000	0.997
7942	91.8%	0.000	0.999
7955	89.3%	0.001	0.990
7964	0.0%	0.000	1.000
7985	100.0%	0.000	1.000
7997	95.7%	0.000	0.998
7998	97.1%	0.001	0.989
8000	93.2%	0.000	0.999
8005	92.4%	0.000	0.996
8018	91.9%	0.000	0.995
8027	3.1%	0.000	0.994
8029	95.8%	0.000	0.998
8030	75.6%	0.004	0.947
8063	83.7%	0.003	0.963
8067	88.7%	0.000	0.996
8079	20.4%	0.001	0.990
8102	98.6%	0.000	0.997
8111	83.2%	0.001	0.985
8119	75.4%	0.001	0.993
8129	71.5%	0.001	0.988
8130	85.3%	0.001	0.990
8131	96.5%	0.000	0.999
8132	94.7%	0.000	0.999
8133	80.4%	0.000	0.997
8134	85.4%	0.000	0.997
8135	64.4%	0.001	0.993
8136	97.0%	0.000	0.998
8142	91.9%	0.001	0.988
8143	91.7%	0.001	0.990
8149	0.0%	0.000	1.000
8160	90.2%	0.000	0.998
8163	10.7%	0.002	0.977
8166	72.7%	0.002	0.973
8167	36.7%	0.008	0.904
8180	92.3%	0.001	0.989

8181	96.7%	0.000	0.995
8199	95.8%	0.000	1.000
8228	83.5%	0.000	0.995
8229	94.0%	0.001	0.991
8242	96.0%	0.000	1.000
8260	73.3%	0.003	0.957
8261	72.6%	0.001	0.987
8262	1.4%	0.000	0.997
8263	90.7%	0.000	0.998
8265	95.8%	0.000	0.999
8277	99.0%	0.000	0.999
8282	40.0%	0.012	0.858
8284	92.0%	0.003	0.961
8288	75.0%	0.023	0.756
8294	100.0%	0.000	1.000
8295	100.0%	0.000	1.000
8300	50.0%	0.125	0.368
8301	50.0%	0.042	0.636
8302	0.0%	0.000	1.000
8305	92.9%	0.005	0.939
8308	60.0%	0.048	0.603
8313	98.8%	0.000	1.000
8320	95.2%	0.002	0.971
8323	100.0%	0.000	1.000
8330	100.0%	0.000	1.000
8332	87.5%	0.014	0.842
8334	100.0%	0.000	1.000
8355	94.1%	0.000	0.997
8369	8.3%	0.000	0.996
8373	97.6%	0.000	0.998
8387	87.7%	0.000	0.995
8396	50.1%	0.000	0.995
8397	85.5%	0.000	0.995
8399	100.0%	0.000	1.000
8401	97.4%	0.000	0.997
8403	94.6%	0.000	0.998
8405	91.4%	0.001	0.985
8407	79.7%	0.001	0.985

8411	97.6%	0.000	1.000
8412	0.0%	0.000	1.000
8414	0.0%	0.000	1.000
8415	83.3%	0.012	0.863
8419	83.5%	0.001	0.993
8421	97.5%	0.000	0.998
8425	72.1%	0.005	0.940
8426	78.2%	0.001	0.985
8427	100.0%	0.000	1.000
8430	72.2%	0.011	0.867
8432	100.0%	0.000	1.000
8438	4.2%	0.002	0.978
8441	67.9%	0.008	0.903
8504	100.0%	0.000	1.000
8506	2.1%	0.000	1.000
8507	90.6%	0.000	0.996
8508	1.3%	0.000	0.999
8510	9.0%	0.000	0.999
8511	9.5%	0.000	0.998
8512	80.6%	0.000	0.996
8513	18.1%	0.002	0.976
8537	66.1%	0.002	0.973
8538	56.0%	0.001	0.990
8542	91.1%	0.000	0.998
8546	95.6%	0.000	0.995
8550	99.5%	0.000	1.000
8551	95.4%	0.000	0.997
8553	96.7%	0.001	0.985
8559	98.6%	0.000	0.997
8561	95.1%	0.000	0.998
8563	83.1%	0.002	0.974
8566	84.2%	0.007	0.912
8568	82.1%	0.004	0.951
8570	88.9%	0.011	0.869
8571	86.8%	0.001	0.989
8573	96.8%	0.001	0.993
8575	94.7%	0.001	0.993
8577	36.4%	0.007	0.912

8579	50.0%	0.042	0.636
8580	82.4%	0.001	0.992
8598	96.0%	0.000	1.000
8608	19.6%	0.003	0.955
8611	72.1%	0.003	0.961
8618	4.2%	0.000	0.997
8624	93.8%	0.000	0.999
8626	100.0%	0.000	1.000
8632	52.2%	0.001	0.989
8635	94.1%	0.003	0.957
8636	80.0%	0.032	0.695
8638	100.0%	0.000	1.000
8639	97.0%	0.001	0.988
8640	100.0%	0.000	1.000
8641	100.0%	0.000	1.000
8644	83.3%	0.023	0.759
8645	100.0%	0.000	1.000
8650	94.7%	0.000	0.996
8651	13.1%	0.001	0.992
8653	100.0%	0.000	1.000

Overall reliability scores (i.e., median of provider-level reliability [R_median], minimum [R_min], maximum [R_max]) by year, and the overall variance between sites, are summarized below.

Overall reliability scores by year, 2010-2014

Year	% suppressed	Var_between	R_median	R_min	R_max
2010	68.4%	0.069	0.990	0.354	1.000
2011	71.1%	0.066	0.991	0.347	1.000
2012	74.3%	0.059	0.991	0.322	1.000
2013	77.5%	0.048	0.991	0.276	1.000
2014	77.6%	0.073	0.996	0.368	1.000

Reliability varied across providers by year. The proportion of providers with reliability greater than or equal to 0.9, 0.8, and 0.7 are shown below.

Distribution of provider-level reliability scores by year, 2010-2014

Year	N	≥0.9 n (%)	≥0.8 n (%)	≥0.7 n (%)
2010	846	793 (93.7)	819 (96.8)	836 (98.8)
2011	811	752 (92.7)	788 (97.2)	792 (97.7)
2012	816	753 (92.3)	788 (96.6)	801 (98.2)
2013	823	753 (91.5)	794 (96.5)	806 (97.9)

2014	813	771 (94.8)	794 (97.7)	802 (98.7)

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

There is no established cut-off for minimum reliability level. Values above 0.7 are considered sufficient to see differences between providers and the mean, and values above 0.9 are considered sufficient to see differences between pairs of providers (RAND Corporation, TR-653-NCQA, 2009).

Each year, more than 91% of providers had reliability scores of 0.9 or greater. Therefore, the reliability of viral suppression can be considered to be sufficient to identify real differences in performance across providers. As previously mentioned, sample size is another driver of reliability and likely contributed to the lowest reliability scores (e.g., in 2014 site 2081 had a reliability of 0.368, and reported 1 of 2 had been prescribed ART). However, median reliability was consistently 0.99 during 2010-2014, supporting the conclusion that the reliability of this measure can be considered very good.

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)*

7. 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

8. Face validity of the performance score was gained through structured presentations (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*)

7. 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.
8. Sixty-nine individuals/organizations submitted 239 pieces of comments. Eight 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?*
7. 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.

8. 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. Eight comments 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)

NA ☒ no exclusions — *skip to section 2b4*

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*)

N/A

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*)

N/A

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*)

N/A

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 ___.

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

☒ **No risk adjustment or stratification**

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

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

☐ **Other,** Click here to enter description

2b4.1.1 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.

The Ryan White HIV/AIDS Program provides a comprehensive system of care that includes primary medical care and essential support services for people living with HIV who are uninsured or underinsured. The Program works with cities, states, and local community-based organizations to provide HIV care and treatment services to more than half a million people each year. The Program reaches approximately 52% of all people diagnosed with HIV in the United States.

As indicated in data presented earlier, the Ryan White HIV/AIDS Program is a public health, safety net program providing care to a high proportion of racial/ethnic minority, transgender, unstable housing, and low income people living with HIV. Many of people served by the Ryan White HIV/AIDS Program represent sociodemographic factors incorporate in risk adjusting models by many measures stewards. As a result, the Ryan White HIV/AIDS Program does not adjust for risk in its performance measures. Rather, it is a fundamental aspect of the Ryan White HIV/AIDS Program to identify disparities and work to improve quality of care for subpopulations. Additionally, this measure is not used for pay-for-performance, bonuses, or penalties.

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)

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

If stratified, skip to ____

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

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

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

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)
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 viral suppression across providers, by year. Performance scores were broken into the bottom 10% and top 90% providers to better characterize the gaps that remain across providers. Moreover, performance scores were examined with respect the proportion of providers with least 80 percent of patients that were prescribed ART in a given year.

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*)

	% patients with viral suppression across providers					providers with ≥80% patients prescribed ART		
Year	Mean	SD	Median	10th %ile	90th %ile	N	n	%
2010	65.9%	27.5%	76.5%	17.8%	91.2%	846	353	41.7
2011	70.1%	26.4%	79.8%	26.1%	93.2%	811	402	49.6
2012	73.4%	25.4%	83.8%	31.7%	94.7%	816	471	57.7
2013	77.5%	24.1%	86.5%	42.9%	96.4%	823	532	64.6
2014	78.0%	28.0%	90.0%	29.6%	98.3%	813	565	69.5

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. In 2014, the bottom 10% of providers had ART prescription rates of 29.6% or lower; the top 90% of providers had rates of 98.3% or higher. These differences demonstrate the continued value of the measure in identifying sites based on poor performance relative to the top performers.

Provider-level performance differences observed in the table above also underscore improvements in the proportion of patients prescribed ART. In 2014, of 813 providers, 565 (69.5%) had prescribed ART for at least 80% of patients. Additionally, on average by provider, nearly 80% (78%) of patients were prescribed ART; however, given the large population that the RWHAP serves, even the poorest performing sites (e.g., bottom 10%) represent a substantial number of patients.

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*)

Based on the method used to calculate the ART performance score, conducting missing data analysis is not applicable for this measure. Specifically, the logic used to determine the number of patients prescribed ART relied on whether or not the patient had at least one medical visit in the measurement year, and then among these patients, whether or not the patient was prescribed ART during the measurement year. Based on provider reporting, patients were classified as either having a medical visit or not, and similarly, patients were considered to be prescribed ART or not, and missing/unknown were not response options.

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*)

N/A (see 2b7.1)

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 (see 2b7.1)

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 collection and availability: The data used for testing and operational use of this measure are readily available within patient health records and provided annually to the Ryan White HIV/AIDS Program through the reporting of the Ryan White Service Report (approved by the Office of Management and Budget 0915-0323).
Missing data: A full analysis of missing data is provided in this submission.
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 Ryan White HIV/AIDS Program https://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio Public Health/Disease Surveillance National HIV/AIDS Strategy https://www.aids.gov/federal-resources/national-hiv-aids-strategy/nhas-update.pdf 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) yan 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

National HIV/AIDS Strategy
Sponsor: Federal government
Geographic area: Nationwide
Accountable entities: Federal agencies and service providers
Patients: All people living with HIV in the United States

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.
Prescription of HIV antiretroviral therapy has been improving in the United States since the first release of publically available data. The Ryan White HIV/AIDS Program served more than 300,000 unduplicated patients annually between 2010-2014 across 2,000+ grant recipients and subrecipients. The Ryan White HIV/AIDS Program has experienced a 10 + point increase in viral suppression from 65.9% in 2010 to 78.0% in 2014. Prescription of HIV antiretroviral therapy has increased across all demographic groups and subpopulations.

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, prescription of HIV antiretroviral therapy is one of five stages of the HIV care continuum. This measure has become the standard when measuring prescription of HIV antiretroviral therapy.

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.
Starting in 2015, Health Resources and Services Administration began releasing December 1st – World AIDS Day – an annual data report (Ryan White HIV/AIDS Program Annual Client-Level Data Report) that contains data similar to those presenting in the report. Building upon the success of the state profiles (http://hab.hrsa.gov/stateprofiles/), Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived). A supplemental report exploring data for the eligible metropolitan areas and transitional grant areas and youth/young adults has been released as well as slides sets for fact sheets by program and population, special populations (http://hab.hrsa.gov/publications/hivaids-bureau-fact-sheets), and infographics (contained in fact sheets). Additionally, grant recipient level reports are prepared and disseminated to all Ryan White HIV/AIDS Program grant recipients.

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.**
Starting in 2015, Health Resources and Services Administration began releasing December 1st – World AIDS Day – an annual data report (Ryan White HIV/AIDS Program Annual Client-Level Data Report) that contains data similar to those presenting in the report. Building upon the success of the state profiles (http://hab.hrsa.gov/stateprofiles/), Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived). A supplemental report exploring data for the eligible metropolitan areas and transitional grant areas and youth/young adults has been released as well as slides sets for fact sheets by program and population, special populations (http://hab.hrsa.gov/publications/hivaids-bureau-fact-sheets), and infographics (contained in fact sheets). Additionally, grant recipient level reports are prepared and disseminated to all Ryan White HIV/AIDS Program grant recipients.

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 from Ryan White HIV/AIDS Program grant recipients and subrecipients regarding the feasibility and usefulness of the data presented in the Ryan White HIV/AIDS Program Annual Client-Level Data Report. Significant feedback has been provided about the timeliness and expansions of the data release. Grant recipient report using the data for benchmarking their program, setting goals/targets, and gaining a fuller understanding of all aspects of the Ryan White HIV/AIDS Program (i.e. other regions of the country). Grant recipients and subrecipients have also requested additional analyses. Health Resources and Services Administration responded with supplemental reports (Ryan White HIV/AIDS Program Supplemental Client-Level Data Report, Eligible Metropolitan Areas and Transitional Grant Areas; special population reports); slide decks for the overall client population and special populations; grant recipient reports; and infographics – all of which will be updated and released annually. Health Resources and Services Administration plans to release additional analyses and special reports this year based on feedback from Ryan White HIV/AIDS Program grant recipients and subrecipients.

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

4d2.3. **Summarize the feedback obtained from other users**
Ryan White HIV/AIDS Program national partners (national organizations that represent grant recipients, subrecipients, and patients) has provided antidotal feedback regarding the timeliness, feasibility, and usability of the release of the Ryan White HIV/AIDS Program Annual Client-Level Data Report, supplemental reports, slide decks, fact sheets, and infographics. The national partners encourage the continued release of the data in all its formats.

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.

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)**

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
2079 HIV Medical Visit Frequency
2080 Gap in HIV Medical Visits

2082 HIV Viral Suppression
3211 Prescription of HIV Antiretroviral Therapy
3210 HIV viral suppression
3010 HIV Medical Visit Frequency
3211 Prescription of HIV Antiretroviral Therapy

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.)
Not applicable.

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.
Employees of hate following governmental and non-governmental organizations/agencies participated in the development of this measure and assisted in assessing face validity:

-HHS Office of HIV/AIDS and Infectious Disease Policy
-Centers for Disease Control
-Center for Medicaid and Medicare
-Health Resources and Services Administration
-Indian Health Service
-National Institutes of Health
-Substances Abuse and Mental Health Services Administration
-U.S. Department of Veterans Affairs
-HIV Medical Association
-Kaiser Permanente
-National Associate of State and Territorial AIDS Directors
-Urban Coalition for HIV/AIDS Prevention Services
-National Minority AIDS Council
-Iowa Department of Health
-Washington D.C. Department of Health
-Maryland Department of Health
-University of Alabama
-University of San Francisco
-Johns Hopkins University

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, 2016

Ad.6 **Copyright statement:** None

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** It is our intention that this measure will be used in quality improvement in addition to public reporting. As it is involved in quality improvement, it is not our intent that the performance goal will be 100%. When we do set the performance goal, we will take into consideration appropriate reasons why the patient may not be able to meet the numerator criterion.

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 #: 3211
Measure Title: Prescription of HIV Antiretroviral Therapy
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 prescribed antiretroviral therapy for the treatment of HIV infection during the measurement year. 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: Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. HIV antiretroviral therapy delays this progression and increases the length of survival.

Current HIV treatment guidelines now recommend universal prescription of HIV antiretroviral therapy for sustained viral load suppression which in turn is directly related to reduction in disease progression and reduction in potential for transmission of HIV infection. Among persons in care, sustained viral load suppression represents the cumulative effect of prescribed therapy, ongoing monitoring, and patient adherence. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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 from the denominator prescribed HIV antiretroviral therapy during the measurement year.
Denominator Statement: Number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the measurement year
Denominator Exclusions: There are no patient exclusions.

Measure Type: Process
Data Source: Electronic Health Record (Only)
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

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**

This measure is the new eMeasure version of NQF #2083. The information provided for Evidence and Opportunity for Improvement is identical to that submitted for NQF #2083. Measure #2083 will be discussed first – the ratings for evidence and opportunity for improvement will automatically be assigned to this eMeasure without further discussion.

Evidence Summary

- 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.
- Evidence and clinical guidelines state that Antiretroviral Therapy is recommended for all HIV-infected individuals in order to reduce morbidity and mortality. Evidence focuses on the percent of providers prescribing ART and the percent of patients with viral load suppression across those providers, the data suggests a positive correlation.
- As a whole, the general evidence suggests that prescription to ART for those infected with HIV will lead to viral suppression if treatment is maintained.
- The developer provided multiple guidelines for the administration of antiretroviral therapy and viral load monitoring intervals for adults, adolescents and pregnant women.
-

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*
  - *For possible exception to the evidence criterion:*
- *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*

Guidance from the Evidence Algorithm

Process measure evidence based (Box3) → Empirical evidence without grading (BOX 7)→ Possible related performance measures (Box 10)→(No exception)→Insufficient

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

RATIONALE: Evidence provided does not directly assess the prescription of ART but rather the viral suppression associated with adherence to ART for HIV infected persons.

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.

Provider-level performance scores for antiretroviral treatment (ART) for 2014 are presented below.

	2014	2013	2012	2011	2010
Rate	77.6	77.5	74.3	71.1	68.4
Pts w/ ≥1 medical visit (den)	316,087	327,618	335,408	327,744	324,455
Pts prescribed ART(num)	245,400 (77.6)	253,972 (77.5)	249,094 (74.3)	233,132 (71.1)	221,908 (68.4)
Mean	78.0	77.5	73.4	70.1	65.9
Median	90.0	86.5	83.8	79.8	76.5
Standard Deviation	28.0	24.1	25.4	26.4	27.5
10th percentile	29.6	42.9	31.7	26.1	17.8
90th percentile	98.3	96.4	94.7	93.2	91.2
Min, Max	0.0, 100.0	0.0, 100.0	0.0, 100.0	0.0, 100.0	0.0, 100.0
# of facilities	813	823	816	811	846

Disparities

- 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. Descriptive characteristics are provided by the developer in the table below. The full table can be found here.

	2014 N (%)	2013 N (%)	2012 N (%)	2011 N (%)	2010 N (%)
Race/Ethnicity					
Am. Indian/Alaska Native	1,272 (0.4)	1,414 (0.5)	1,371 (0.4)	1,366 (0.4)	1,473 (0.5)
Asian	3,791 (1.2)	3,835 (1.2)	3,980 (1.2)	3,598 (1.2)	3,382 (1.1)
Black/African Am.	142,746 (46.9)	146,056 (47.0)	150,974 (47.2)	149,834 (47.8)	146,460 (47.3)
Hispanic/Latino	74,714 (24.5)	74,967 (24.1)	75,201 (23.5)	71,240 (22.7)	71,002 (22.9)
Native Hawaiian/Pacific Is.	442 (0.2)	510 (0.2)	575 (0.2)	710 (0.2)	627 (0.2)
White	75,931 (24.9)	78,953 (25.4)	83,820 (26.2)	83,061 (26.5)	83,854 (27.1)
Multiple Races	5,651 (1.9)	4,899 (1.6)	4,238 (1.3)	3,716 (1.2)	3,177 (1.0)
Gender					
Male	216,965 (70.7)	221,930 (70.7)	230,075 (70.8)	223,379 (69.9)	219,625 (69.7)
Female	87,071 (28.4)	89,212 (28.4)	92,186 (28.4)	93,687 (29.3)	93,266 (29.6)
Transgender	2,974 (1.0)	2,779 (0.9)	2,848 (0.9)	2,585 (0.8)	2,313 (0.7)

Questions for the Committee:

- *Without data from the eMeasure as specified, do you agree that there is a quality problem with prespcription of HIV antiretroviral therapy?*
- *Are you aware of evidence that other disparities exist in prescribing HIV antiretroviral therapy?*

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)

1a. Evidence
*Evidence consistent with the NQF #2083 Measure = INSUFFICIENT

1b. Performance Gap
*Same as the NQF #2083 = MODERATE

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):
- Electronic Health Records (only). This is an eMeasure

Specifications:
- HQMF specifications for the eMeasure are included in the document set on SharePoint. See eMeasure Technical Advisor review below.
- This measure is specified level of analysis is at hospital/facility/agency level
- Patients are included in the numerator if they were prescribed HIV antiretroviral therapy during the measurement year
- The denominator includes the number of patients, regardless of age, with a diagnosis of HIV with at least one medical visit in the measurement year
- There are no patient exclusions
- The calculation algorithm 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.
- The value sets needed to calculate the numerator and denominator are included in the specifications.

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?*

eMeasure Technical Advisor(s) review (if not an eMeasure, delete this section):

Submitted measure is an	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)).

HQMF compliant eMeasure	**HQMF specifications ☒ Yes ☐ No**
Documentation of HQMF or QDM limitations	**N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM**
Value Sets	**The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC**
Measure logic is unambiguous	**Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously**
Feasibility Testing	**The submission contains a feasibility assessment that addresses data element feasibility and follow-up with measure developer indicates that the measure logic is feasible based on assessment by EHR vendors**

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.

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

- The dataset used for testing included 34 synthetic patients created in the Bonnie testing system simulating the year 2012. The developer tested the following data elements using the Bonnie testing tool to evaluate the measure logic:
  - Patient name
  - Date of birth
  - Race
  - Ethnicity
  - Gender
  - Payer
  - Diagnosis
  - Medication Orders
  - Encounters
- The patient bundle's demographics were designed to mimic the HIV/AIDS population, specifically drawing from the patient characteristics collected via the Ryan White HIV/AIDS Program Services Report (RSR).
- Data element validity testing was performed and will count for data element reliability – see validity testing section.
- The developer provided reliability results from the chart-abstracted measure (#2083) and stated, "Currently, there is no performance data available to test the eCQM. However, the chart-abstracted version of this measure has been in use in national quality reporting programs since as early as 2010."

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*

- *Do the results from the Bonnie tool demonstrate sufficient reliability so that differences in performance can be identified?*
- *Do you agree that the reliability test results of the eMeasure will be comparable to the paper based measure (#2083)?*

Guidance from the Reliability Algorithm Precise specifications (Box 1) →Empirical reliability testing (Box 2) →Empirical validity testing of patient-level data (Box 3) →Refer to validity testing of patient-level data elements using Bonnie tool (Box 10 of the Validity algorithm) →Method appropriate for legacy eMeasures (Box 11) → Moderate is the highest possible rating

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:

- *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.

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

Validity testing method:

- The Bonnie testing tool, with 34 synthetic patient records were used to test the measure logic and data elements.
  - For each synthetic patients, an expected result was assigned to reflect an expected result of the measure. The synthetic patients were then run against the HQMF output loaded into Bonnie, which "calculates" a measure result for each patients and evaluates it against the expected result.
  - A patient is considered to pass Bonnie testing when the expected result matches the "calculated" result.
- The following testing was completed on the synthetic patients
  - 100% logic coverage: The bundle of synthetic patients collectively includes all data elements and conditions that are specified within the measure logic.
  - Clinical relevance. References cited within the chart abstracted measure specification were used to design clinically relevant, realistic patient profiles for the measure's target population. This approach ensured the eCQM logic maintained alignment with the clinical intent of the chart abstracted measure.
  - Edge case testing: Data elements that test the upper or lower boundary of measure logic conditions.
  - Negative testing: Use of test cases that do not evaluate positively against the measure logic but are otherwise clinically relevant and realistic.
- The developer used references cited within the chart abstracted measure specifications to ensure the eCQM logic maintained alignment with the clinical intent of the chart abstracted measure.

- In addition to Bonnie testing, the measure specifications were reviewed independently by three eCQM experts to confirm the logic was syntactically correct, using appropriate and current versions of the eCQM standards and terminologies, and consistent with the intent of the chart-abstracted measure.

Validity testing results:

- The testing results from the Bonnie tool reached 100% coverage and confirmed there was a test case for each pathway of logic (negative and positive test cases).
- The measure had a 100% passing rate which confirmed that all the test cases performed as expected.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

No exclusions

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)*:

The Data represents variability across providers, In 2014, the bottom 10% of providers had ART prescription rates of 29.6% or lower; the top 90% of providers had rates of 98.3% or higher. These differences demonstrate the continued value of the measure in identifying sites based on poor performance relative to the top performers.

% patients with viral suppression across providers							Providers with ≥80% patients prescribed ART		
Year	Mean	SD	Median	10th %ile	90th %ile		N	n	%
2010	65.9%	27.5%	76.5%	17.8%	91.2%		846	353	41.7
2011	70.1%	26.4%	79.8%	26.1%	93.2%		811	402	49.6
2012	73.4%	25.4%	83.8%	31.7%	94.7%		816	471	57.7
2013	77.5%	24.1%	86.5%	42.9%	96.4%		823	532	64.6
2014	78.0%	28.0%	90.0%	29.6%	98.3%		813	565	69.5

Question for the Committee:

- *Does the Committee agree the e-Measure will demonstrate similar results to the chart-abstracted measure?*

2b6. Comparability of data sources/methods:

- Not applicable

2b7. Missing Data

- Per the developer, "The HQMF standard specifies that if data are unknown or missing, they shall fail the criterion. This constraint embodies the notion that absence of evidence is evidence of absence, i.e. data not present in a structured field from which the measure draws will not be considered for measure calculation. In certain cases, missing data may have no impact on the measure outcome for a given patient. For example, a data element used in a series of OR statements will not impact the measure outcome if another data element in the OR statement is present and meets all other defined constraints."
- All Bonnie synthetic patients with missing data performed according to the HQMF standard specification and as expected.

Guidance from the Validity Algorithm Specifications consistent with evidence (Box 1)→Some threats to validity addressed (Box 2) →Empirical validity testing (Box 3) → Face validity testing (Box 4) and empirical testing of data elements using Bonnie tool (Box 10) →Method appropriate for legacy eMeasures (Box 11) →Moderate is the highest possible rating

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

Committee pre-evaluation comments

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

2a1. Reliability Specifications
*All data elements are well defined. All codes are provided, while no descriptors are available other than the field definition All steps are clear. No concerns that this can be consistently implemented. - Due to Field testing and no Score testing, Reliability = MODERATE

*The data elements are clearly defined
All appropriate codes included
The logic or calculation algorithm is clear
It is likely this measure can be consistently implemented"

2a2. Reliability Testing
*Reliability testing good. Measure is a legacy eCQM. Elements are "yes"/"no" responses based on data in the medical record and provide good reliability. Not a PRO-PM measure.

*The reliability test results of the eMeasure should be comparable to the paper based measure (#2083)

2b1. Validity Specifications
*None. Not a PRO-PM measure. No empirical validity testing. Validity = MODERATE

2b2. Validity Testing
*Adequate testing. Bonnie Testing Tool used which is a gold standard for data element testing. No empirical validity testing. Validity = MODERATE.

*The test sample is adequate to generalize for widespread implementation
The results demonstrate sufficient validity so that conclusions about quality can be made
It is not clear that medications prescribed are received by the patient."

2b3-7 Threats to Validity
*Missing data/responses are not considered in calculation of the rates. "absence of evidence is evidence of absent".

*2b.5 The e-Measure will demonstrate similar results to the chart-abstracted measure
"

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 provided information on feasibility testing in the eMeasure Feasibility Score Card . The developer did not identify the EHRs used for feasibility testing. Instead, the developer stated that the feasibility assessment was "conducted by consensus of a panel of MITRE clinical informatics, measure development, and eCQM standards experts".
- The developer provided a summary of the latest publicly available data on Meaningful Use EHR capabilities and provider performance on objectives and measures related to the eCQM's data elements:
  - CPOE – Meds
  - CPOE – Labs
  - Demographics
  - Problem List
  - Lab test results
- On a scale from 1 to 3 where 3 is the highest score, all but 3 of the data elements received a score of '3'.
  - Both 'Encounter, Performed: Face to Face Interaction' and 'Patient Characteristic Payer' scored a 2 on Data Standards.
    - The Score 2 definition for Data Standards is "terminology standards for this data element are currently available, but it is not consistently coded to standard terminology in the EHR, or the EHR does not easily allow such coding."
  - The data element 'Patient Characteristic Expired' scored a 2 on Data Accuracy. Data accuracy looks at the correctness of the information contained in the data element and whether the data source and recorder are specified.
    - The Score 2 definition for Data Accuracy is "the information may not be from the most authoritative source and/or has a moderate likelihood of being correct". The scorecard notes that this information is similar to "self-reporting of a vaccination".
    - The developer notes that "The accuracy of this data element is dependent on full end-to-end interoperability across providers and between providers and public health agencies."
- The developer indicates that on a scale from 0 to 100 percent, the measure is currently **98.33%** feasible and in one to two years, will be **98.89%** feasible.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Is the data collection strategy ready to be put into operational use?*
- *Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility
*The required data elements are routinely generated/used in clinical care delivery. The data collection strategy is already in place. No concerns.

*The required data elements are routinely generated and used during care delivery
The data collection strategy is ready to be put into operational use
I was unable to access the eMeasure Feasibility Score Card "

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**

OR

Planned use in an accountability program? ☐ **Yes** ☒ **No**

Accountability program details

- 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 (PQRS) and Value Based Modifier
  - Sponsor: Federal government
  - Geographic area: Nationwide
  - Accountable entities: Physicians and practitioners
  - Patients: Unknown
- National HIV/AIDS Strategy
  - Sponsor: Federal government
  - Geographic area: Nationwide
  - Accountable entities: Federal agencies and service providers
  - Patients: All people living with HIV in the United States

Improvement results

- The developer reports that the Ryan White HIV/AIDS Program has experienced a 10 + point increase in viral suppression from 65.9% in 2010 to 78.0% in 2014. Prescription of HIV antiretroviral therapy has increased across all demographic groups and subpopulations.

Unexpected findings (positive or negative) during implementation

- 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, prescription of HIV antiretroviral therapy is one of five stages of the HIV care continuum. This measure has become the standard when measuring prescription of HIV antiretroviral therapy.

Potential harms

- Not applicable

Vetting of the measure

- The developer reports that Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected which is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports).
- Ryan White HIV/AIDS Program national partners has provided feedback regarding the timeliness, feasibility, and usability of the release of the Ryan White HIV/AIDS Program Annual Client-Level Data Report, supplemental reports, slide decks, fact sheets, and infographics. The national partners encourage the continued release of the data in all its formats.
- During the initial development of the chart-abstracted measure, formal feedback was gathered. The measures were modified during the development phase and have not been modified since. On an annual basis, the

measures are reviewed for clinical relevance, change in scientific acceptability, and consistency with guidelines. The chart-abstracted measure has not been modified as a result of the annual reviews.

Feedback:

- The developer reports that feedback has been received from Ryan White HIV/AIDS Program grant recipients and subrecipients regarding the feasibility and usefulness of the data presented in the Ryan White HIV/AIDS Program Annual Client-Level Data Report. Significant feedback has been provided about the timeliness and expansions of the data release. Grant recipient report using the data for benchmarking their program, setting goals/targets, and gaining a fuller understanding of all aspects of the Ryan White HIV/AIDS Program. Grant recipients and subrecipients have also requested additional analyses.

***Questions for the Committee**:*

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

*Data are currently reported publically and used to identify providers who need improvement in this area. It is also being used in a pay-4-performance programs. The results are also used as benchmarks for quality improvement work being done in clinics both inside and outside of the Ryan White HIV/AIDS Program.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis
- 0409 HIV/AIDS: Sexually Transmitted Disease Screening for Chlamydia, Gonorrhea, and Syphilis
- 2079 HIV Medical Visit Frequency
- 2080 Gap in HIV Medical Visits
- 2082 HIV Viral Suppression
- 3210 HIV Viral Suppression
- 3010 HIV Medical Visit Frequency

Harmonization

Measure is fully harmonized to the extent possible according to the developer

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 candidate 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**

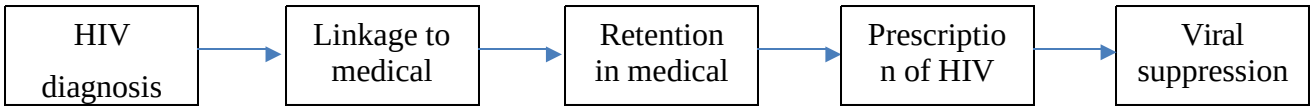
RATIONALE IF NOT ELIGIBLE: The measure is not eligible for Endorsement+ because empirical reliability and validity testing of the measure score was not conducted and the measure has not been vetted in real world settings by those being measured and other users.

Pre-meeting public and member comments

-

Measure Title: Prescription of HIV Antiretroviral Therapy

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.

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)

Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. HIV antiretroviral therapy delays this progression and increases the length of survival.

Current HIV treatment guidelines now recommend universal prescription of HIV antiretroviral therapy for sustained viral load suppression which in turn is directly related to reduction in disease progression and reduction in potential for transmission of HIV infection. Among persons in care, sustained viral load suppression represents the cumulative effect of prescribed therapy, ongoing monitoring, and patient adherence. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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.

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 15, 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 15, 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 15, 2016. http://www.iapac.org/uploads/JIAPAC-IAPAC-Guidelines-for-Optimizing-the-HIV-Care-Continuum-Supplement-Nov-Dec-2015.pdf

Günthard HF, Saag MS, Benson CA, del Rio C, Eron JJ, Gallant JE, Hoy JF, Mugavero MJ, Sax PE, Thompson MA, Gandhi RT, Landovitz RJ, Smith DM, Jacobsen DM, Volberding PA. Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults 2016 Recommendations of the International Antiviral Society–USA Panel. JAMA. 2016. https://www.iasusa.org/content/antiretroviral-drugs-treatment-and-prevention-hiv-infection-adults-2016-recommendations

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

Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents:

Initiation of Antiretroviral Therapy (page E-1)

- Antiretroviral therapy (ART) is recommended for all HIV-infected individuals, regardless of CD4 T lymphocyte cell count, to reduce the morbidity and mortality associated with HIV infection (AI).
- ART is also recommended for HIV-infected individuals to prevent HIV transmission (AI).
- When initiating ART, it is important to educate patients regarding the benefits and considerations regarding ART, and to address strategies to optimize adherence. On a case-by-case basis, ART may be deferred because of clinical and/or psychosocial factors, but therapy should be initiated as soon as possible.

Considerations for Antiretroviral Use in Special Patient Populations: Acute and Recent (Early) HIV Infection (page I-1)

- Antiretroviral therapy (ART) is recommended for all individuals with HIV-1 infection (AI) including those with early HIV-1 infection.

HIV-Infected Adolescents and Young Adults (page I-8):

- ART is recommended for all HIV-infected individuals (AI) to reduce morbidity and mortality. Thus, ART is also recommended for ART-naive adolescents. However, before initiation of therapy, adolescents' readiness and ability to adhere to therapy within their psychosocial context need to be carefully considered as partner of therapeutic decision making (AIII).

HIV-Infected Women (page I-20):

- Antiretroviral therapy (ART) is recommended for all HIV-infected women to improve their health and to reduce the risk of HIV transmission to HIV-uninfected sex partners (AI).

HIV/Hepatitis C Virus Coinfection (page J-6):

- Antiretroviral therapy (ART) may slow the progression of liver disease by preserving or restoring immune function and reducing HIV related immune activation and inflammation. For most HCV/HIV-coinfected

patients, including those with cirrhosis, the benefits of ART outweigh concerns regarding drug-induced liver injury. Therefore, ART should be initiated in all HCV/HIV-coinfected patients, regardless of CD4 T lymphocyte (CD4) cell count (AI).

WHO:

4.3 When to start ART (page xxxi)

4.3.1 When to start ART in adults (>19 years old)

- ART should be initiated in all adults living with HIV, regardless of WHO clinical stage and at any CD4 cell count (strong recommendation, moderate-quality evidence).
- As a priority, ART should be initiated in all adults with severe or advanced HIV clinical disease (WHO clinical stage 3 or 4) and adults with a CD4 count ≤350 cells/mm3 (strong recommendation, moderate-quality evidence).

4.3.2 When to start ART in pregnant and breastfeeding women

- ART should be initiated in all pregnant and breastfeeding women living with HIV, regardless of WHO clinical stage and at any CD4 cell count and continued lifelong (strong recommendation, moderate-quality evidence).

4.3.3 When to start HIV antiretroviral therapy in adolescents (10–19 years of age)

- ART should be initiated in all adolescents living with HIV, regardless of WHO clinical stage and at any CD4 cell count (conditional recommendation, low-quality evidence).
- As a priority, ART should be initiated in all adolescents with severe or advanced HIV clinical disease (WHO clinical stage 3 or 4) and adolescents with a CD4 count ≤350 cells/mm3 (strong recommendation, moderate-quality evidence).

4.3.4 When to start HIV antiretroviral therapy in children younger than 10 years of age

- ART should be initiated in all children living with HIV, regardless of WHO clinical stage or at any CD4 cell count:
- Infants diagnosed in the first year of life (strong recommendation, moderate-quality evidence).
- Children living with HIV 1-year-old to less than 10 years old (conditional recommendation, low-quality evidence).
- As a priority, ART should be initiated in all children <2 years of age or children younger than 5 years of age with WHO clinical stage 3 or 4 or CD4 count ≤750 cells/mm³ or CD4 percentage <25% and children 5 years of age and older with WHO clinical stage 3 or 4 or CD4 count ≤350 cells/mm³ (strong recommendation, moderate-quality evidence).

4.3.5 Timing of HIV ANTIRETROVIRAL THERAPY for adults and children with TB

- ART should be started in all TB patients living with HIV regardless of CD4 count (strong recommendation, high-quality evidence).

International Advisory Panel on HIV Care Continuum Optimization (IAPAC):

Increasing HIV treatment coverage (page 3)

- The immediate offer of ART after HIV diagnosis, irrespective of CD4 count or clinical stage, is recommended. (AI)

Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults 2016 Recommendations of the International Antiviral Society–USA Panel

Box 1. Recommendations for When to Start (page 193)

- Antiretroviral therapy (HIV ANTIRETROVIRAL THERAPY) is recommended for all viremic patients with established HIV infection, regardless of CD4 cell count (evidence rating AIa).
- Initiation of ART is recommended as soon as possible in the setting of acute HIV infection (evidence rating BIII).
- Planned discontinuation of early ART after a specific duration of treatment is not recommended outside a research setting (evidence rating AIa).
- Initiation of ART is recommended for individuals who have persistent undetectable viral load without ART but have declining CD4 cell counts (evidence rating BIII).

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:

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

Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults, 2016 Recommendations of the International Antiviral Society–USA Panel:

Table 1. Strength of Recommendation and Quality of Evidence Rating Scale

Rating	Definition
Strength of recommendation	
A	Strong support for the recommendation
B	Moderate support for the recommendation
C	Limited support for the recommendation
Quality of evidence	
Ia	Evidence for ≥ 1 randomized clinical trials published in the peer-reviewed literature
Ib	Evidence for ≥ 1 randomized clinical trials presented in abstract form at peer-reviewed scientific meetings
IIa	Evidence from nonrandomized clinical trials or cohorts or case-control studies published in the peer-reviewed literature
IIb	Evidence from nonrandomized clinical trials or cohorts or case-control studies published in the peer-reviewed scientific meeting
III	Recommendation based on panel's analysis of the accumulated available evidnce

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

ART_evidence_NQF-636174955634964398-636177547774061167.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.

Human immunodeficiency virus (HIV) is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 50,000 persons in the United States are newly infected with HIV each year. Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection. HIV antiretroviral therapy delays this progression and increases the length of survival.

Current HIV treatment guidelines now recommend universal prescription of HIV antiretroviral therapy for sustained viral load suppression which in turn is directly related to reduction in disease progression and reduction in potential for transmission of HIV infection. Among persons in care, sustained viral load suppression represents the cumulative effect of prescribed therapy, ongoing monitoring, and patient adherence. The proposed measure will direct providers' attention and quality improvement efforts towards this important outcome.

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 "ART 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 "ART 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.)*
There is no measure-specific web page for the electronic version of this measure.

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 an eMeasure **Attachment:** NQFXXX_PrescriptionOfAntiretroviralTherapy_Artifacts-636177547766417118.zip,NQFXXX_PrescriptionOfAntiretroviralTherapy_MeasureSubmissionForm-636177547766573119.docx

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:** HIVPAT_v4_6_Thu_Dec_15_20.34.08_CST_2016.xls

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.

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 from the denominator prescribed HIV antiretroviral therapy during the measurement year.

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).
The antiretroviral therapy medication order is represented by the QDM element “Medication, Order: FDA Approved HIV Antiretroviral Therapy” using “HIV Antiretroviral Therapy RXNORM Value Set (2.16.840.1.113762.1.4.1032.1).” In order to be included in the numerator, the “Medication, Order: FDA Approved HIV Antiretroviral Therapy” element must start during the measurement 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 measurement year

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).
The patient’s HIV diagnosis is represented by the QDM element "Diagnosis: HIV" using "HIV Grouping Value Set (2.16.840.1.113883.3.464.1003.120.12.1003)".

The patient's medical visits are represented by the following QDM elements:

- "Diagnosis: HIV 1" using "HIV 1 Grouping Value Set (2.16.840.1.113883.3.464.1003.120.12.1004)"
- "Encounter, Performed: Face-to-Face Interaction" using "Face-to-Face Interaction Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1048)"
- "Encounter, Performed: Office Visit" using "Office Visit Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1001)"
- "Encounter, Performed: Outpatient Consultation" using "Outpatient Consultation Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1008)"
- "Encounter, Performed: Preventive Care - Established Office Visit, 0 to 17" using "Preventive Care - Established Office Visit, 0 to 17 Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1024)"
- "Encounter, Performed: Preventive Care Services - Established Office Visit, 18 and Up" using "Preventive Care Services - Established Office Visit, 18 and Up Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1025)"
- "Encounter, Performed: Preventive Care Services-Initial Office Visit, 18 and Up" using "Preventive Care Services-Initial Office Visit, 18 and Up Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1023)"
- "Encounter, Performed: Preventive Care- Initial Office Visit, 0 to 17" using "Preventive Care- Initial Office Visit, 0 to 17 Grouping Value Set (2.16.840.1.113883.3.464.1003.101.12.1022)"

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
There are no patient exclusions.

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.)*
There are no patient exclusions.

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.)*
N/A

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 measurement year or prior to the measurement year; and 2.) had at least one medical visit during the measurement year. The individuals who met these criteria are the denominator population.
2. Identify the individuals from the denominator population who meet the criterion for inclusion in the numerator: prescribed HIV antiretroviral therapy during the measurement year.
3. Calculate the percentage 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.
Electronic Health Record (Only)

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.
Not applicable.

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
ART_testing-636177547781081212.docx,NQFXXX_PrescriptionOfAntiretroviralTherapy_BonnieTestingAttachment-636177547781237213.zip

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.)
Yes

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*)**:** Click here to enter NQF number
Measure Title: Prescription of HIV Antiretroviral Therapy
Date of Submission: 12/16/2016
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is

precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the

measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Synthetic Bonnie test patients	☒ other: Synthetic Bonnie test patients

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*).

This measure is a legacy electronic clinical quality measure (eCQM) – an NQF endorsed measure currently used in federal quality programs that has been respecified into eMeasure. Per NQF modified testing requirements for legacy eCQMs, the measure was tested in the Bonnie testing tool. Bonnie is designed to validate eCQM specifications (HQMF output and value sets) against the measure's expected behavior for user-developed synthetic test patients.

The synthetic patient bundle used to test this measure was designed to simulate clinically relevant, realistic patient scenarios aligned with the target population for this measure. Full details on the Bonnie synthetic patient bundle used to test this measure are included in the Bonnie testing attachment.

For more information on Bonnie, please visit https://bonnie.healthit.gov/.

1.3. What are the dates of the data used in testing? The Bonnie test environment simulates the year 2012 as the measurement period.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

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

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*)

Not applicable. The Bonnie synthetic patient bundle was used to test the measure.

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*)

A test bundle of 34 patients was designed and built within the Bonnie testing tool to evaluate the measure logic. Information documented for each patient within the bundle include:
Patient name
Date of birth
Race
Ethnicity
Gender
Payer

Additional elements contained within the patient profiles as appropriate for testing against expected outcomes include:
Diagnosis
Medication orders
Encounters

The patient bundle's demographics were designed to mimic the HIV/AIDS population, specifically drawing from the patient characteristics collected via the Ryan White HIV/AIDS Program Services Report (RSR).

The breakdown of test bundle demographics for the 34 patients included (represented by number of patients/percentage of bundle): males 23/68%; females 11/32%; American Indian/Alaska Native 1/3%; Asian 1/3%; Black/African American 15/44%; Native Hawaiian/Pacific Islander 0/0%; White 9/26%; Hispanic/Latino

8/24%; younger than 13 1/3%; 13-17 years old 1/3%; 18-24 years old 2/6%; 25-34 years old 6/18%; 35-44 years old 6/18%; 45-54 years old 10/29%; 55-65 years old 6/18%; older than 65 2/6%.
Full details on the Bonnie synthetic patient bundle used to test this measure, including human-readable and QRDA Category 1 format documents for each synthetic patient record, are included in the Bonnie testing attachment.

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.

The Bonnie patient test deck was used to satisfy all testing requirements for this measure. The testing results are further supported by testing data for the chart-abstracted version of this measure collected through the Health Resources and Services Administration HIV/AIDs Bureau's Ryan White HIV/AIDS Program Services Report.

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).

Patient sociodemographic variables considered in the analysis of the chart-abstracted version of this measure were included in the eCQM specifications and modeled in the Bonnie patient bundle. These variables included age, race, ethnicity, gender and payer.

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

Currently, there is no performance data available to test the eCQM. However, the chart-abstracted version of this measure has been in use in national quality reporting programs since as early as 2010.

The most recent reliability analysis of the chart-abstracted measure 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.

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*)

Overall reliability scores (i.e., median of provider-level reliability [R_median], minimum [R_min], maximum [R_max]) by year, and the overall variance between sites, are summarized below.

Table 1. Overall reliability scores by year, 2010-2014

Year	% suppressed	Var_between	R_median	R_min	R_max
2010	68.4%	0.069	0.990	0.354	1.000
2011	71.1%	0.066	0.991	0.347	1.000
2012	74.3%	0.059	0.991	0.322	1.000
2013	77.5%	0.048	0.991	0.276	1.000
2014	77.6%	0.073	0.996	0.368	1.000

Reliability varied across providers by year. The proportion of providers with reliability greater than or equal to 0.9, 0.8, and 0.7 are shown below.

Table 2. Distribution of provider-level reliability scores by year, 2010-2014

Year	N	≥0.9 n (%)	≥0.8 n (%)	≥0.7 n (%)
2010	846	793 (93.7)	819 (96.8)	836 (98.8)
2011	811	752 (92.7)	788 (97.2)	792 (97.7)
2012	816	753 (92.3)	788 (96.6)	801 (98.2)
2013	823	753 (91.5)	794 (96.5)	806 (97.9)
2014	813	771 (94.8)	794 (97.7)	802 (98.7)

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

There is no established cut-off for minimum reliability level. Values above 0.7 are considered sufficient to see differences between providers and the mean, and values above 0.9 are considered sufficient to see differences between pairs of providers (RAND Corporation, TR-653-NCQA, 2009).
Each year, more than 91% of providers had reliability scores of 0.9 or greater. Therefore, the reliability of viral supression can be considered to be sufficient to identify real differences in performance across providers. As previously mentioned, sample size is another driver of reliability and likely contributed to the lowest reliability scores (e.g., in 2014 site 2081 had a reliability of 0.368, and reported 1 of 2 had been prescribed ART). However, median reliability was consistently 0.99 during 2010-2014, supporting the conclusion that the reliability of this measure can be considered very good.

2b2. VALIDITY TESTING

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

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

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

The Bonnie testing environment was used to test the validity of the measure logic and data elements. For each Bonnie synthetic patient, an expected measure result was assigned to reflect the expected outcome of the measure given the specific patient scenario and associated data. The synthetic patients were run against the HQMF output loaded into Bonnie, which produces a measure outcome for each patient and evaluates it against the expected outcome. A patient is considered to pass Bonnie testing when the expected outcome matches the actual outcome, e.g. when a patient is expected to be in the numerator population and the computation of the synthetic patient data against the eCQM logic places the patient in the numerator. In order to achieve a rigorous, clinically relevant test bundle, synthetic patients were designed following the below principles and test areas:

- Clinical relevance. References cited within the chart abstracted measure specification were used to design clinically relevant, realistic patient profiles for the measure's target population. This approach ensured the eCQM logic maintained alignment with the clinical intent of the chart abstracted measure.
- 100% logic coverage: The resulting bundle of synthetic patients collectively includes all data elements and conditions logic that are specified within the measure logic, including at least one patient evaluating against each measure population pathway. Fully testing the measure logic increases test rigor and mitigates risk of unexpected outcomes.
- Edge case testing. Edge cases refer to those data elements that test the upper or lower boundary of measure logic conditions, e.g. a diagnosis starting on the latest qualifying date or a medication order for antiretroviral therapy starting on the first day or last day of the measurement period. Edge cases are designed to test each edge that exists within each measure population.
- Negative testing. Negative testing involves use of test cases do not evaluate positively against measure logic, but are otherwise clinically relevant and realistic, e.g. scenarios where an HIV diagnosis was not documented or a medication order for antiretroviral therapy starting on the first day or last day of the measurement period. Negative testing further validates measure logic by accurately evaluating patients against expected outcomes and simulating the effect of missing data on measure results.

In addition to Bonnie testing, the measure specifications were reviewed independently by three eCQM experts to confirm the logic was syntactically correct, using appropriate and current versions of the eCQM standards and terminologies, and consistent with the intent of the chart-abstracted measure.

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

Bonnie testing results provide logic coverage and passing rates. The synthetic bundle reached 100% coverage, confirming each logic pathway was tested. The results also showed 100% passing rate, confirming all synthetic patients performed as expected.

Full details on Bonnie testing results are contained in the Bonnie testing attachment. The attachment includes a human-readable (HTML) summary document that lists each patient within the bundle and its passing status against expected measure outcomes. The attachment also includes a summary spreadsheet for the synthetic patient bundle which lists each patient, associated demographics, expected and actual measure population outcomes, and which portions or each measure population logic the patient meets expectations for.

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?*)

The results of measure logic testing through use of Bonnie provided confidence in the measure logic accurately representing the clinical intent and alignment with the chart abstracted measure.

2b3. EXCLUSIONS ANALYSIS (FOR MEASURS WITH EXCLUSIONS --- gap in visits and medical visit frequency)
NA ☒ no exclusions — *skip to section ___*

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*)

Not applicable.

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*)

Not applicable.

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*)

Not applicable.

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 ___.

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

2b4.1.1 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.

Not applicable.

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.

The Ryan White HIV/AIDS Program provides a comprehensive system of care that includes primary medical care and essential support services for people living with HIV who are uninsured or underinsured. The Program works with cities, states, and local community-based organizations to provide HIV care and treatment services

to more than half a million people each year. The Program reaches approximately 52% of all people diagnosed with HIV in the United States.

As indicated in data presented earlier, the Ryan White HIV/AIDS Program is a public health, safety net program providing care to a high proportion of racial/ethnic minority, transgender, unstable housing, and low income people living with HIV. Many of people served by the Ryan White HIV/AIDS Program represent sociodemographics factors incorporate in risk adjusting models by many measures stewards.

As a result, the Ryan White HIV/AIDS Program does not adjust for risk in its performance measures. Rather, it is a fundamental aspect of the Ryan White HIV/AIDS Program to identify disparities and work to improve quality of care for subpopulations. Additionally, this measure is not used for pay-for-performance, bonuses, or penalties.

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*)

Not applicable.

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

Not applicable.

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)

Not applicable.

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

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

Not applicable.

If stratified, skip to _____

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:** Not applicable.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*): Not applicable.

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves: Not applicable.

2b4.9. Results of Risk Stratification Analysis: Not applicable.

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*)

Not applicable.

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

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

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

The chart-abstracted version of this measure has been in use since 2010. To examine meaningful differences in performance, we examined the distribution of the proportion of patients with viral suppression across providers, by year. Performance scores were broken into the bottom 10% and top 90% providers to better characterize the gaps that remain across providers. Moreover, performance scores were examined with respect the proportion of providers with least 80 percent of patients that were prescribed ART in a given year.

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*)

	% patients with viral suppression across providers					providers with ≥80% patients prescribed ART		
Year	Mean	SD	Median	10th %ile	90th %ile	N	n	%
2010	65.9%	27.5%	76.5%	17.8%	91.2%	846	353	41.7
2011	70.1%	26.4%	79.8%	26.1%	93.2%	811	402	49.6
2012	73.4%	25.4%	83.8%	31.7%	94.7%	816	471	57.7
2013	77.5%	24.1%	86.5%	42.9%	96.4%	823	532	64.6
2014	78.0%	28.0%	90.0%	29.6%	98.3%	813	565	69.5

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 variablility 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. In 2014, the bottom 10% of providers had ART prescription rates of 29.6% or lower; the top 90% of providers had rates of 98.3% or higher. These differences demonstrate the continued value of the measure in identifying sites based on poor performance relative to the top performers.

Provider-level performance differences observed in the table above also underscore improvements in the proportion of patients prescribed ART. In 2014, of 813 providers, 565 (69.5%) had prescribed ART for at least 80% of patients. Additionally, on average by provider, nearly 80% (78%) of patients were prescribed ART;

however, given the large population that the RWHAP serves, even the poorest performing sites (e.g., bottom 10%) represent a substantial number of patients.

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*)

Not applicable

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*)

Not applicable

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*)

Not applicable

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*)

The HQMF standard specifies that if data are unknown or missing, they shall fail the criterion. This constraint embodies the notion that absence of evidence is evidence of absence, i.e. data not present in a structured field from which the measure draws will not be considered for measure calculation. In certain cases, missing data may have no impact on the measure outcome for a given patient. For example, a data element used in a series of OR statements will not impact the measure outcome if another data element in the OR statement is present and meets all other defined constraints.

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)

The Bonnie synthetic patient bundle includes scenarios for missing data elements, which are a form of negative testing. All Bonnie synthetic patients with missing data performed according to the HQMF standard specification and as expected.

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*)

Please see response for question 2b7.1 above.

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).
For this measure, we are presenting an e-measure and paper measure.

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: NQFXXX_PrescriptionOfAntiretroviralTherapy_Feasibility_Scorecard_v1.0-636177547788569260.xlsx,NQFXXX_PrescriptionOfAntiretroviralTherapy_MeasureTestingAttatchment-636177547788725261.docx

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.
Not applicable.

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)*.
The measure specifications contain limited proprietary codes for convenience. Users of CPT(R) should obtain all necessary licenses from the owners of these code sets.
The use of SNOMED Clinical Terms(R) requires a Unified Medical Language System (UMLS) license. These licenses are freely available, from the National Library of Medicine.

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 Ryan White HIV/AIDS Program https://hab.hrsa.gov/clinical-quality-management/performance-measure-portfolio Public Health/Disease Surveillance National HIV/AIDS Strategy https://www.aids.gov/federal-resources/national-hiv-aids-strategy/nhas-update.pdf 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) yan 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

National HIV/AIDS Strategy
Sponsor: Federal government
Geographic area: Nationwide
Accountable entities: Federal agencies and service providers
Patients: All people living with HIV in the United States

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.
Prescription of HIV antiretroviral therapy has been improving in the United States since the first release of publically available data. The Ryan White HIV/AIDS Program served more than 300,000 unduplicated patients annually between 2010-2014 across 2,000+ grant recipients and subrecipients. The Ryan White HIV/AIDS Program has experienced a 10 + point increase in viral suppression from 65.9% in 2010 to 78.0% in 2014. Prescription of HIV antiretroviral therapy has increased across all demographic groups and subpopulations.

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, prescription of HIV antiretroviral therapy is one of five stages of the HIV care continuum. This measure has become the standard when measuring prescription of HIV antiretroviral therapy.

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.
Starting in 2015, Health Resources and Services Administration began releasing December 1st – World AIDS Day – an annual data report (Ryan White HIV/AIDS Program Annual Client-Level Data Report) that contains data similar to those presenting in the report. Building upon the success of the state profiles (http://hab.hrsa.gov/stateprofiles/), Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived). A supplemental report exploring data for the eligible metropolitan areas and transitional grant areas and youth/young adults has been released as well as slides sets for fact sheets by program and population, special populations (http://hab.hrsa.gov/publications/hivaids-bureau-fact-sheets), and infographics (contained in fact sheets). Additionally, grant recipient level reports are prepared and disseminated to all Ryan White HIV/AIDS Program grant recipients.

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.**
Starting in 2015, Health Resources and Services Administration began releasing December 1st – World AIDS Day – an annual data report (Ryan White HIV/AIDS Program Annual Client-Level Data Report) that contains data similar to those presenting in the report. Building upon the success of the state profiles (http://hab.hrsa.gov/stateprofiles/), Health Resources and Services Administration worked diligently to release the annual data report in the same year it was collected (collected in April and released in December of the same year). The report is publically available on the Health Resources and Services Administration website (http://hab.hrsa.gov/data/data-reports) and is released via an accompanying webinar (recorded and archived). A supplemental report exploring data for the eligible metropolitan areas and transitional grant areas and youth/young adults has been released as well as slides sets for fact sheets by program and population, special populations (http://hab.hrsa.gov/publications/hivaids-bureau-fact-sheets), and infographics (contained in fact sheets). Additionally, grant recipient level reports are prepared and disseminated to all Ryan White HIV/AIDS Program grant recipients.

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 from Ryan White HIV/AIDS Program grant recipients and subrecipients regarding the feasibility and usefulness of the data presented in the Ryan White HIV/AIDS Program Annual Client-Level Data Report. Significant feedback has been provided about the timeliness and expansions of the data release. Grant recipient report using the data for benchmarking their program, setting goals/targets, and gaining a fuller understanding of all aspects of the Ryan White HIV/AIDS Program (i.e. other regions of the country). Grant recipients and subrecipients have also requested additional analyses. Health Resources and Services Administration responded with supplemental reports (Ryan White HIV/AIDS Program Supplemental Client-Level Data Report, Eligible Metropolitan Areas and Transitional Grant Areas; special population reports); slide decks for the overall client population and special populations; grant recipient reports; and infographics – all of which will be updated and released annually. Health Resources and Services Administration plans to release additional analyses and special reports this year based on feedback from Ryan White HIV/AIDS Program grant recipients and subrecipients.

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

4d2.3. **Summarize the feedback obtained from other users**
Ryan White HIV/AIDS Program national partners (national organizations that represent grant recipients, subrecipients, and patients) has provided antidotal feedback regarding the timeliness, feasibility, and usability of the release of the Ryan White HIV/AIDS Program Annual Client-Level Data Report, supplemental reports, slide decks, fact sheets, and infographics. The national partners encourage the continued release of the data in all its formats.

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.

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)**

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
2079 HIV Medical Visit Frequency
2080 Gap in HIV Medical Visits
2082 HIV Viral Suppression
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.**

This measure does not have a competing measure.

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.)

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.

Employees of hate following governmental and non-governmental organizations/agencies participated in the development of this measure and assisted in assessing face validity:

-HHS Office of HIV/AIDS and Infectious Disease Policy

-Centers for Disease Control

-Center for Medicaid and Medicare

-Health Resources and Services Administration

-Indian Health Service

-National Institutes of Health
-Substances Abuse and Mental Health Services Administration
-U.S. Department of Veterans Affairs
-HIV Medical Association
-Kaiser Permanente
-National Associate of State and Territorial AIDS Directors
-Urban Coalition for HIV/AIDS Prevention Services
-National Minority AIDS Council
-Iowa Department of Health
-Washington D.C. Department of Health
-Maryland Department of Health
-University of Alabama
-University of San Francisco
-Johns Hopkins University

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, 2016

Ad.6 **Copyright statement:** None

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** It is our intention that this measure will be used in quality improvement in addition to public reporting. As it is involved in quality improvement, it is not our intent that the performance goal will be 100%. When we do set the performance goal, we will take into consideration appropriate reasons why the patient may not be able to meet the numerator criterion.

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 #: 3215
Measure Title: Adult Inpatient Risk Adjusted Sepsis Mortality
Measure Steward: New York State Department of Health, Office of Quality and Patient Safety
Brief Description of Measure: Annual risk adjusted inpatient mortality rate for adult patients (aged 18 and over) admitted to acute care hospitals with diagnosis of severe sepsis or septic shock. The measure includes patients in acute care hospital settings over one year timeframe who had, either on admission, or during their hospital stay, a clinical diagnosis of severe sepsis (now referred to as ´sepsis´) or septic shock using criteria described in the International Sepsis Definitions (Sepsis-2)

Hospitals were required to submit a protocol for early identification and treatment of severe sepsis or septic shock. Subsequent to protocol submission, hospitals were required to submit 100% of their patient cases to a data collection portal using a standardized data dictionary (see relevant sections for details). Numerous data elements including patient demographics and comorbidities among other patient care details were reported. A random sample of the data submissions were validated for accuracy. The full adult data for discharges within calendar year 2015 was used to generate statewide and hospital-specific risk adjusted mortality rates for the calendar year.
Developer Rationale: Mortality is an important outcome for patients with sepsis. Mortality rates are high and show significant variability across acute care hospitals unrelated to patient factors. International, national, and local system improvement demonstrations have shown considerable reductions in mortality with focused efforts at early diagnosis and timely treatment. In order to use mortality as an outcome measure, a robust sepsis-specific risk model is essential.

Numerator Statement: Outcome is risk adjusted inpatient mortality rate for adult patients (18 and over) admitted to an acute care hospital with a diagnosis of severe sepsis or septic shock or who develop severe sepsis or septic shock during their hospital stay.
Denominator Statement: All adult patient discharges (18 and over) in a calendar year with a diagnosis of severe sepsis or septic shock on admission or at any time during their hospital stay. This may include multiple admissions of the same patient during the measurement year. Denominator includes all cases identified using any means (administrative, registry, electronic health records, billing data, etc.), either prospectively, retrospectively, or both, that meet the International consensus definition (Sepsis- 2) of severe sepsis or septic shock.
Denominator Exclusions: Patients with advanced directives in place prior to episode of sepsis which specifically restrict any hospital specific sepsis protocol interventions or who decline (or their proxy declines) treatment for sepsis. Patients who have been transferred from one acute care hospital to another are excluded.

Measure Type: Outcome
Data Source: Claims (Other), EHRs Hybrid, Laboratory, Management Data, Non-Medical Data, Paper Records, Pharmacy, Registry
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary

- Mortality is an important outcome for patients with sepsis and according to the developer, mortality rates are high and show significant variability across acute care hospitals unrelated to patient factors. The developer reports that international, national, and local system improvement demonstrations have shown reductions in mortality with focused efforts at early diagnosis and timely treatment.
- As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence mortality rates through the use of early sepsis detection approaches coupled with rapid delivery of basic resuscitation interventions including use of adequate intravenous fluids, antibiotics, blood pressure support medications and dynamic clinical monitoring for response.

Question for the Committee:

- *Does the Committee agree that at least one hospital process identified by the developer impacts inpatient sepsis mortality?*

Guidance from the Evidence Algorithm: Health outcome measure (Box 1)→The relationship between the outcome and at least one process is identified and supported by the stated rationale (Box 2)→Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

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

- The developer provided the risk-adjusted probability of inpatient sepsis mortality rates from 179 hospitals and 43,204 patients in New York State from January 1, 2015 – December 31, 2015 for this newly developed measure:

Table 1. Probablity of hospital mortality by year and quarter

Year and Quarter	Mean	SD	Min	1st quartile	Median	3rd quartile	Max
Q1 2015	30.4	19.6	0.1	14.4	27.0	44.0	93.2
Q2 2015	28.9	19.6	0.1	12.7	25.1	42.1	94.8
Q3 2015	28.8	19.7	0.1	12.5	25.3	42.2	93.8
Q4 2015	28.4	20.0	0.1	11.8	24.4	42.0	95.0
Total	29.1	19.7	0.1	12.8	25.4	42.6	95.0

Table 2. Decile of the probability of hospital mortality by year and quarter

Decile of probability	Q1 2015	Q2 2015	Q3 2015	Q4 2015	Total
1	4.1	3.3	3.1	2.9	3.3
2	9.5	8.3	7.9	7.6	8.3
3	14.4	12.7	12.5	11.8	12.8
4	19.2	17.3	17.3	16.6	17.6
5	24.2	22.3	22.5	21.8	22.7
6	30.1	28.3	28.2	27.5	28.5
7	36.5	34.9	34.7	34.4	35.1
8	44.0	42.3	42.3	42.1	42.7

9	53.1	51.6	51.6	51.4	51.9
10	68.6	67.7	67.8	68.3	68.1

Disparities:

- The developer provided the following probability of inpatient sepsis mortality rates by population group:

Table 1. Probability of mortality by population group

	Probability of hospital mortality
Race/Ethnicity	
White, non-Hispanic	28.2
Black, non-Hispanic	31.5
Hispanic	26.3
Multi-racial	29.5
Unknown, non-Hispanic	30.7
Unknown	32.0
Gender	
Female	28.4
Male	29.8
Age Categories	
18-29	10.3
30-39	15.3
40-49	20.4
50-59	25.0
60-69	28.7
70-79	31.0
80+	33.9
Insurance/payer	
Medicare	30.6
Medicaid	26.2
Private, HMO	27.1
Self-Pay	34.3
Other	26.1

Questions for the Committee:

- *Does the data demonstrate variation and less-than-optimal probability of inpatient sepsis mortality rates?*
- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that other disparities exist in this area of healthcare?*

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)

1a. Evidence:

- This measure just requires reporting of mortality which is a PRO but isn't tied to a specific process or intervention
- I agree that at least one hospital process identified by the developer impacts inpatient sepsis mortality
- Yes, strong and well substantiated relationship

1b. Performance Gap

- Yes
- the data demonstrate variation and less-than-optimal probability of inpatient sepsis mortality rates
- There is a gap in care that warrants a national performance measure
- Subgroups are identified - disparities are present but limited

1c. Composite Performance Measure - Quality Construct (if applicable): Are the following stated and logical: overall quality construct, component performance

- yes

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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): Claims (Other), EHRs Hybrid, Laboratory, Management Data, Non-Medical Data, Paper Records, Pharmacy, Registry. This is not an eCQM.

Specifications:

- The level of analysis is facility-level.
- The numerator is the risk adjusted inpatient mortality rate for adult patients (18 and over) admitted to an acute care hospital with a diagnosis of severe sepsis or septic shock or who develop severe sepsis or septic shock during their hospital stay
- The denominator includes all adult patient discharges (18 and over) in a calendar year with a diagnosis of severe sepsis or septic shock on admission or at any time during their hospital stay.
  - This may include multiple admissions of the same patient during the measurement year.
  - Denominator includes all cases identified using any means (administrative, registry, electronic health records, billing data, etc.), either prospectively, retrospectively, or both, that meet the International consensus definition (Sepsis- 2) of severe sepsis or septic shock.
- The denominator exclusions include patients:
  - Advanced Directives in place prior to diagnosis of severe sepsis or septic shock that specifically preclude active treatment according to that hospital´s protocol for severe sepsis and septic shock.
  - Patient or patient proxy refusal of treatment for severe sepsis or septic shock according to that hospital´s protocol for severe sepsis and septic shock.
  - Patients who were transferred between acute care hospitals.
- This outcome measure is risk-adjusted using a multivariate logistic regression model.
- Better quality is associated with a lower score.
- The calculation algorithm describes how the risk-adjusted sepsis mortality rate is calculated.
- A standardized clinical data dictionary with set specified data fields is included.

Questions for the Committee:

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

2a2. Reliability Testing Testing attachment

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.

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**

- Reliability testing performed at the facility level of analysis; however, the developer is encouraged to clarify the data source.

Method(s) of reliability testing: Data element validity testing was performed and will count for data element reliability as well - see validity testing section.

Guidance from the Reliability Algorithm: Precise specifications (Box 1) → Empirical reliability testing conducted using statistical tests the measure as specified (Box 2) → Measure score reliability testing was not conducted (Box 4) → Abtraction of patient-level data elements compared to an authoritative source/gold standard – see validity algorithm (Box 8) → Patient-level data element testing conducted (Box 10) → Compared original abstraction from facilities to the gold-standard (external auditors). Only assessed percent agreement; did not provide additional statistics such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) (Box 11) → Insufficient

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

RATIONALE: Patient-level data element validity testing is insufficient because developer only assessed percent agreement.

2b. Validity

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**

- This outcome measure calculates the risk-adjusted sepsis mortality rate for adults admitted to acute care hospitals with a diagnosis of severe sepsis or septic shock.
- The rationale for measuring this health outcome suggests that hospitals are able to influence mortality rates through the use of early sepsis detection approaches coupled with rapid delivery of basic resuscitation interventions including use of adequate intravenous fluids, antibiotics, blood pressure support medications and dynamic clinical monitoring for response.

Question for the Committee:

- *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.

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:

- The dataset included hospitals in New York State that were required to develop and implement early recognition and treatment protocols for sepsis. As part of this statewide initiative, hospitals were required to submit quarterly clinical data to the New York State Department of Health to be evaluated for protocol use, adherence to time interventions and patient outcomes, including mortality.
    - The dataset included a total of 179 hospitals with 43,204 patients diagnosed with severe sepsis and septic shock from January 1, 2015 to December 31, 2015.
    - The dataset used to develop the logistic regression model included:
        - Development sample: 38,884 (90%) patients; 179 hospitals
        - Validation sample: 4,319 (10%) patients; 160 hospitals

Data element validity testing

- The developer validated the accuracy of the data submission from the hospitals against manual chart abstraction by external auditors (Audit Results), which is considered the gold standard. The developer calculated the percent agreement between the hospital submissions to the gold standard. *In addition to percent agreement, NQF criteria requires sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV).*
    - The developer also validated the accuracy of the external auditors' ability to abstract the hospital data from the medical record (IRR POA). Auditors were required to reach 100% accuracy prior to validating hospital submissions.

Measure score validity testing

- Empirical validity testing of the measure score was assessed by comparing the performance of the risk-adjusted model in the development sample to the validation sample.
    - The developer used the Hosmer-Lemeshow goodness of fit test to assess the observed and expected mortality rates in the development and validation samples.
        - The development dataset was split into group sizes of 10, 100, 500 and 1,000. The p-values were 0.568, 0.972, 0.735, and 0.735, respectively.
        - The validation dataset was split into group sizes of 10, 50, 100 and 150. The p-values were 0.651, 0.977, 0.985, and 0.974, respectively.

Validity testing results:

Data element results

- The table below includes the percent agreement between the data submission from the hospitals and the gold standard (external auditors). Percent agreement from the audit results ranged from 89.9% to 99.1% for the data elements below - sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) not provided.
- Per the developer, the following data elements were not audited manually but aligned to state administrative datasets to ensure accuracy: race, ethnicity, payer, admission source.

Variable Name	Audit Results (% Validity)	IRR Gold Standard	IRR POA	IRR Kappa
Race	N/A	N/A	N/A	N/A
Ethnicity	N/A	N/A	N/A	N/A
Payer	N/A	N/A	N/A	N/A
Site of Infection	98.9%	100.0%	100.00	SC
Admission Source	N/A	100.0%	93.33	0.6341
Lower Respiratory Infection	98.8%	100.0%	93.33	SC
Mechanical Ventilation	97.8%	100.0%	100	SC
Age (Date of Birth)	99.2%	100.0%	96.67	SC
Thrombocytopenia	97.7%	100.0%	93.33	SC
Septic Shock	98.4%	100.0%	100.00	SC
Serum Lactate (Lactate Level)	93.9%	100.0%	96.67	0.8519
Metastatic Cancer	97.1%	100.0%	100.00	1
Lymphoma, Leukemia, Multiple Myeloma	99.1%	100.0%	96.67	SC
Square Root of Comorbidity Count (Range of Comorbidities)	89.9%-99.1%	100.0%	96.67-100.00	SC
*N/A=These variables were not audited manually but were aligned to state administrative datasets to ensure accuracy. SC=There were insufficient cell spread to enable kappa calculation. This can happen with high percent agreement with the same response category.				

Measure score results

- The performance of the risk-adjustment model was similar in the development and validation datasets. The areas under the receiver operating characteristic (ROC) curve (or *c*-statistic) were **0.770** and **0.773**, respectively.
  - A *c*-statistic is a model of discrimination statistic. A *c*-statistic of 0.77 means that 77% of all possible pairs of patients – one who died and one who lived – the model correctly assigned a higher probability to those who died. Generally, a *c*-statistic of at least 0.70 is considered acceptable. The similar *c*-statistics indicates good model discrimination.
- The developer also provided a table with the adjusted hospital mortality and the adjusted odds ratio.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the methods and results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- For the development of the risk model, the developer stated that they excluded a patient's prior admissions and kept the final hospital admission for those patients with multiple admissions to New York State hospitals during the testing period because:
  - The outcome needed to be statistically independent - a necessary assumption when running a logistic regression model.
  - It would be impossible for a patient with multiple hospital admissions for sepsis to experience hospital mortality any but their last admission.
- Patients with advanced directives in place that precluded one or more elements of the sepsis protocol were also excluded. Additionally, if the patient or surrogate decision maker declined interventions they were excluded.
- The developer stated that they did not test the exclusions for the following reasons:
  - Keeping prior hospital admissions in the dataset violates the principle of statistical independence of the outcome.

- Keeping patients in the dataset that had an advanced directive or declined intervention would not be appropriate since the population of interest are those that were eligible to receive the treatment.
- The developer noted that keeping those excluded patients in the analysis would bias the results because:
  - Keeping prior hospital admissions in the dataset would bias the results towards lower hospital mortality since it is impossible to experience hospital mortality since they obviously survived in order to be admitted to hospital (i.e. their last observation).
  - Keeping patients in the dataset that had an advanced directive or declined intervention would bias the results towards higher hospital mortality.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

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

Conceptual rationale for SDS factors included ? ☐ **Yes** ☒ **No**

SDS factors included in risk model? ☒ **Yes** ☐ **No**

Risk adjustment summary

Description of the model

- This measure is risk-adjusted using a multivariable logistic regression model with 16 variables to estimate the probability of mortality for patients admitted to acute care hospitals with severe sepsis or septic shock.
- The model was built using the development dataset and starting with all possible covariates in the model. Using an iterative procedure, variables were removed from the model, one by one, if the *p*-values were not significant at 0.05 level until a parsimonious model was reached.
  - Variables removed during the development procedure were added back into the reduced model if the p-values were significant at the 0.05 level and if model calibration (Hosmer-Lemeshow goodness of fit) was improved through their inclusion.
  - The scale of the 3 continuous variables (patient age, first serum lactate, and the count of the number of comorbidities) remaining in the model was assessed. Using the method of fractional polynomials patient age was included in the model as a linear term, number of comorbidities was transformed by taking the square root, and first serum lactate was entered into the model as a quadratic expression (linear and a squared term).
  - Model calibration was further improved by adding the following interactions to the model: lower respiratory infection (LRI) by MV severity, patient age by the square root of the number of comorbidities, and first serum lactate by the square root of the number of comorbidities.

Performance of the model

Discrimination statistics:

- The areas under the receiver operating characteristic (ROC) curve (or *c*-statistic) reflects how accurately a statistical model is able to distinguish between a patient with an outcome and a patient without an outcome. C-statistic values can range from 0.5 to 1.0. A value of 0.5 indicates that the model is no better than chance at making a prediction of patients with and without the outcome of interest and a value of 1.0 indicates that the model perfectly identifies those with and without the outcome of interest . Generally, a c-statistic of at least 0.70 is considered acceptable.
- The *c*-statistic value was computed using the dataset randomly split into two samples:
  - The *c*-statistic for the development sample was 0.770
  - The *c*-statistic for the validation sample was 0.773
- A *c*-statistic of 0.77 means that 77% of all possible pairs of patients – one who died and one who lived –the model correctly assigned a higher probability to those who died. Generally, a *c*-statistic of at least 0.70 is considered acceptable. The similar *c*-statistics indicates good model discrimination.

SDS factors in risk-adjustment approach

- The following demographic variables were included for model building though not all demographic variables remained viable in the final model: age; gender; payor; race; and, ethnicity.
- Inclusion in the model was based on whether or not the demographic variable was statistically significant at the 0.05 level.
- Gender was the only variable included in the model since its odds ratio and corresponding p-value was 1.0003 and 0.992, respectively. All of the other demographic variables had p-values < 0.001 for at least one of the levels of a specific demographic.

Questions for the Committee:

- *Is an appropriate risk-adjustment strategy included in the measure?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*

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

- The developer provided performance data.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- The patient's first serum lactate was the only variable used in the model with missing data where 6.0% were missing in those that survived their hospital stay and 5.6% were missing in those that died in the hospital.
- Single imputation using truncated linear regression was used during the imputation procedure where the lower limit of left truncation was set at a serum lactate level 0.1 mmol/L (1st percentile) and the upper limit of the right truncation was set at 30.0 mmol/L (99th percentile).
- The developer provided the table of 24 predictor variables used to impute the missing serum lactate.
- When the risk adjusted model was run using the imputed serum lactate the overall risk-adjusted hospital mortality was 29.1%.
- When the observations associated with missing lactate were dropped from the model development the overall risk adjusted hospital mortality was 29.4%.
- The developers concluded that statistical methods can be used to impute the missing serum lactate.

Guidance from the Validity Algorithm: Specifications consistent with the provided in support of the measure (Box 1) → Potential threats to validity that are relevant to the mesure empirically assessed (Box 2) → Empirical validity testing conducted using the measure as specified (Box 3) → Validity testing conducted with computed performance measure scores (Box 6) → Method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships (Box 7) → Moderate certainty or confidence that the performance measure scores are a valid indicator of quality (8b) → Moderate

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

Committee pre-evaluation comments

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

2a1. & 2b1. Specifications: Reliability-Specifications - Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not

- Case mix is definitely something that could skew data as there are very different levels of severity within "severe sepsis"

- the data elements are clearly defined All appropriate codes included The calculation algorithm is clear It is likely this measure can be consistently implemented

2a1. & 2b1. Specifications: Reliability-Specifications:

- Case mix is definitely something that could skew data as there are very different levels of severity within "severe sepsis
- the data elements are clearly defined, All appropriate codes included, The calculation algorithm is clear. It is likely this measure can be consistently implemented

2b.1 Validity – Specifications:

- None
- The specifications are consistent with the evidence
- Validity testing in generally on target

2b2. Validity - Testing:

- The test sample is adequate to generalize for widespread implementation
- NY-state centric but generalizable

2b3-7. Exclusions, Risk Adjustment, Statistically Significant Differences, Multiple Data Sources, Missing Data

- **2b.3** I don't understand why excluding patients with multiple admissions is appropriate. It seems this would artificially increase the mortality rate.

**Tiering of hospitals based on their case-mix would be useful

- **2b.4** an appropriate risk-adjustment strategy is included in the measure

The candidate and final variables included in the risk adjustment model are adequately described for the measure to be implemented
serum lactate levels may vary depending on when the samples drawn in relation to the course of the septic episode.

- **2b.5** This measure identifies meaningful differences about quality

2a2. Reliability - Testing:
*Yes

Criterion 3. Feasibility

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, coded by someone other than person obtaining original information, and abstracted from a record by someone other than person obtaining original information.
- The developer notes that some data elements are in defined fields in electronic sources, and that some demographic variables can be extracted electronically and used in a standard format. Other variables are collected manually by hospitals though some hospitals have created electronic data capture avenues.
- The developer notes that "it is more likely that hospitals belonging to a larger system of hospitals have been able to general more electronic data capture opportunities. All of the data is submitted following

standardized data element definitions, file structure, type and requirements with algorithms to ensure data complies with known logic criteria".

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?*
- *Is the data collection strategy ready to be put into operational use?*
- *Are there fees or licenses associated with the use of this measure?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility:

- None
- The required data elements may not be routinely generated and used during care delivery

Criterion 4: Usability and Use

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

Public Reporting

- Program/Sponsor: New York State Sepsis Improvement Initiative; New York State Department of Health
- Purpose: Transparency and accountability with respect to outcomes of treatment for patients with severe sepsis/septic shock in acute care facilities in NY; quality improvement and recognition of promising practices in care delivery
- Geography/Entities/Patients: Statewide; ~180 facilities; ~50,000 patients/year
- Level of Measurement/Setting: population based risk adjusted inpatient sepsis mortality within acute care settings

Quality Improvement

- Program/Sponsor: New York State Sepsis Improvement Initiative; New York State Department of Health
- Purpose: Quality improvement to assist in the recognition of both positive and negative outlier performance, identification of promising practices, analysis of interventions (protocols, bundles, etc.) on patients including subpopulations; used in conjunction with Partnership for Patients care improvement with New York hospitals and New York HENs (hospital associations)
- Geography/Entities/Patients: Statewide; ~180 facilities; ~50,000 patients/year
- Level of Measurement/Setting: risk adjusted inpatient sepsis mortality within acute care settings

Regulatory Programs

- Program/Sponsor: New York State Sepsis Improvement Initiative; New York State Department of Health

- Purpose: fulfills Title 10 Regulation (Sections 405.2 and 405.4) in New York State and Section 2805-m of Public Health Law requiring the collection and reporting by the Department of data regarding the performance of hospitals for patients with sepsis including risk adjusted mortality rates for individual hospitals
- Geography/Entities/Patients: Statewide; ~180 institutions; ~50,000 patients
- Level of Measurement/Setting: risk adjusted inpatient sepsis mortality rates within acute care settings

Improvement results

- The developer notes that preliminary evidence shows both raw and risk adjusted mortality improvements over time, using data from 2Q 2014 to 3Q 2016. The developer reports that data from hospital performance during CY2015 will be released to the public in early 2017.

Unexpected findings (positive or negative) during implementation

- N/A

Potential harms

- N/A

Vetting of the measure

- The developer reports that the New York Sepsis Advisory Group provides input into the development of variables in the data dictionary for purposes of eventual risk adjustment. The group provides ongoing input into the data dictionary, reviewed an intial risk adjustment model and supplied comments and suggestions which were incorporated into the final model.
- Hospitals were involved in decisions regarding patient exclusions.

Feedback:

- The developer receives feedback through in person meetings, phone conference and webinars with the advisory group. Hospitals and clinicials also provide feedback and comments via email, phone, and letters.
- The developer reports that feedback on the risk adjustment model, variables, and performance has been positive and that some specialized hospitals were concerned about whether the model adequately adjusts or considers their particular patient population and risks.
- Feedback on data collection suggests that abstraction remains a significant task/burden for hospitals.
- Users identified the need to account for patient mix differences in evaluation of mortality since risk adjustment allows users to identify those institutions for outreach who have demonstrated outstanding results, understand the impact of interventions, hospital characteristics associated with better or worse outcomes and so on.
- Several variables in the initial model were revised based on feedback from the advisory group.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*
- *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:

- I think this could be useful, but only if case-mix is applied to the numbers
- The measure (or similar) has been vetted and used by others

Criterion 5: Related and Competing Measures
Related or competing measures • 0500: Severe Sepsis and Septic Shock: Management Bundle **Harmonization** • Measure #3215 is newly submitted to the Infectious Disease project. The related and competing discussion for these measures will take place during the in person meeting in March. • The developer reports the measure specifications are harmonized to the extent possible.

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 candidate 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**: Reliability of the measure score not conducted.

Pre-meeting public and member comments

•

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 3215
Measure Title: Inpatient Sepsis Risk Adjusted Mortality Rate
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: 1/31/2017

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*

- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: Risk Adjusted Sepsis Inpatient Mortality Rate

☐ Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Sepsis Mortality is an outcome of importance to both clinicians and patients. Risk adjusted mortality rates permit fair comparisons of care quality between health care organizations reducing the probability that differences are a result

of patient mix differences, versus care process differences. There are a number of clinical interventions that are known to impact this outcome, making risk adjusted sepsis mortality rates an 'actionable' outcome measure.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

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).

Reductions in sepsis mortality has been associated with use of early sepsis detection approaches coupled with rapid delivery of basic resuscitation interventions including use of adequate intravenous fluids, antibiotics, blood pressure support medications and dynamic clinical monitoring for response.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	

Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

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

NQF_evidence_attachment_FINAL_2016.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.

No

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.

Mortality is an important outcome for patients with sepsis. Mortality rates are high and show significant variability across acute care hospitals unrelated to patient factors. International, national, and local system improvement demonstrations have shown considerable reductions in mortality with focused efforts at early diagnosis and timely treatment. In order to use mortality as an outcome measure, a robust sepsis-specific risk model is essential.

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.*

Table 1: Descriptive stats of the probability of hospital mortality by year and quarter

Year and quarter	Mean	SD	Min	1st Q	Median	3rd Q	Max
2015 1	30.4	19.6	0.1	14.4	27.0	44.0	93.2
2015 2	28.9	19.6	0.1	12.7	25.1	42.1	94.8
2015 3	28.8	19.7	0.1	12.5	25.3	42.2	93.8
2015 4	28.4	20.0	0.1	11.8	24.4	42.0	95.0
Total	29.1	19.7	0.1	12.8	25.4	42.6	95.0

Table 2: Decile of the probability of hospital mortality by year and quarter

Decile of probability	Year and quarter				Total
	2015 1st	2015 2nd		2015 3rd	2015 4th
1	4.1	3.3	3.1	2.9	3.3
2	9.5	8.3	7.9	7.6	8.3
3	14.4	12.7	12.5	11.8	12.8
4	19.2	17.3	17.3	16.6	17.6
5	24.2	22.3	22.5	21.8	22.7
6	30.1	28.3	28.2	27.5	28.5
7	36.5	34.9	34.7	34.4	35.1
8	44.0	42.3	42.3	42.1	42.7
9	53.1	51.6	51.6	51.4	51.9
10	68.6	67.7	67.8	68.3	68.1

Description of data

Collected from 179 New York State hospitals from January 1, 2015 to December 31, 2015

Number of patients in the analysis: 43,204

Number of variables collected by New York State: 73

Number of variables used in the prediction model: 16 – listed below in Table 3

Table 3: List of variables used in the logistic regression model

Race/ethnicity

White, non-Hispanic (Referent)

Black, non-Hispanic

Hispanic

Multi-racial

Unknown, non-Hispanic

Unknown

Payer

Medicare (Referent)

Medicaid

Private, HMO

Self-Pay

Other
Site of infection
Urinary (Referent)
Respiratory
Gastrointestinal
Skin
Central nervous system
Other
Unknown
Admission source
Non-health facility, POA (Referent)
Clinic
Different Hospital
Skilled nursing facility/Intermediate care facility
Another health care facility
Between unit transfer
Hospice
Other
Lower respiratory infection
No (Referent)
Yes
MV severity
No (Referent)
Yes
Lower respiratory infection and MV severity - Interaction
Septic shock diagnosis
Severe sepsis
Septic Shock
Thrombocytopenia
No (Referent)
Yes
Metastatic cancer
No (Referent)
Yes
Lymphoma/leukemia/multiple myeloma
No (Referent)
Yes
Age
Square root of comorbidity count
Age and the square root of the comorbidity count - Interaction
Serum lactate
Serum lactate squared
Serum lactate and the square root of the comorbidity count - Interaction

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.**

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.*

Table 1: Probability of mortality by population group

Population group	Probability of hospital mortality
Race/Ethnicity	
White, non-Hispanic	28.2
Black, non-Hispanic	31.5
Hispanic	26.3
Multi-racial	29.5
Unknown, non-Hispanic	30.7
Unknown	32.0
Gender	
Female	28.4
Male	29.8
Age categories	
18-29	10.3
30-39	15.3
40-49	20.4
50-59	25.0
60-69	28.7
70-79	31.0
80+	33.9
Insurance/payer	
Medicare	30.6
Medicaid	26.2
Private, HMO	27.1
Self-Pay	34.3
Other	26.1

Description of data

Collected from 179 New York State hospitals from January 1, 2015 to December 31, 2015

Number of patients in the analysis: 43,204

Number of variables collected by New York State: 73

Number of variables used in the prediction model: 16 – listed below in Table 3

Table 3: List of variables used in the logistic regression model

Race/ethnicity
White, non-Hispanic (Referent)
Black, non-Hispanic
Hispanic
Multi-racial
Unknown, non-Hispanic
Unknown
Payer
Medicare (Referent)
Medicaid
Private, HMO
Self-Pay
Other
Site of infection
Urinary (Referent)
Respiratory
Gastrointestinal
Skin
Central nervous system
Other

Unknown
Admission source
Non-health facility, POA (Referent)
Clinic
Different Hospital
Skilled nursing facility/Intermediate care facility
Another health care facility
Between unit transfer
Hospice
Other
Lower respiratory infection
No (Referent)
Yes
MV severity
No (Referent)
Yes
Lower respiratory infection and MV severity - Interaction
Septic shock diagnosis
Severe sepsis
Septic Shock
Thrombocytopenia
No (Referent)
Yes
Metastatic cancer
No (Referent)
Yes
Lymphoma/leukemia/multiple myeloma
No (Referent)
Yes
Age
Square root of comorbidity count
Age and the square root of the comorbidity count - Interaction
Serum lactate
Serum lactate squared
Serum lactate and the square root of the comorbidity count - Interaction

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**

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):*

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):*

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.)*

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:** Sepsis_Data_Dictionary_3.0_pub-636214687710592961.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.

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).
Outcome is risk adjusted inpatient mortality rate for adult patients (18 and over) admitted to an acute care hospital with a diagnosis of severe sepsis or septic shock or who develop severe sepsis or septic shock during their hospital stay.

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).
Inpatient mortality is noted on data submission from hospital. Clinical variables needed for risk adjustment including demographics, co-morbidities, severity, and potential exclusions are reported by hospital as described in the data dictionary.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All adult patient discharges (18 and over) in a calendar year with a diagnosis of severe sepsis or septic shock on admission or at any time during their hospital stay. This may include multiple admissions of the same patient during the measurement year.

Denominator includes all cases identified using any means (administrative, registry, electronic health records, billing data, etc.), either prospectively, retrospectively, or both, that meet the International consensus definition (Sepsis- 2) of severe sepsis or septic shock.

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).
All adult patients meeting International consensus definition (Sepsis-2) for Severe Sepsis/Septic shock identified through combination of any relevant hospital clinical and/or administrative databases, prospectively or retrospectively.

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients with advanced directives in place prior to episode of sepsis which specifically restrict any hospital specific sepsis protocol interventions or who decline (or their proxy declines) treatment for sepsis. Patients who have been transferred from one acute care hospital to another are excluded.

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 who have any of the following characteristics, reported on data variables fully described in the data dictionary, are excluded from the calculation of risk adjusted mortality rates for a specific hospital:
1. Advanced Directives in place prior to diagnosis of severe sepsis or septic shock that specifically preclude active treatment according to that hospital´s protocol for severe sepsis and septic shock.
2. Patient or patient proxy refusal of treatment for severe sepsis or septic shock according to that hospital´s protocol for severe sepsis and septic shock.
3. Patients who were transferred between acute care hospitals.

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.)*
Stratification
The analysis was not stratified for different populations since there was only a single population studied: patients with sepsis. However in the risk adjusted logistic regression model there are categorical variables that represent either patient demographics or patient clinical characteristics. This mix of variables generates the probability of mortality across the levels of the categorical variable. For example septic shock diagnosis is in the model so a probability of hospital mortality could be generated for both severe sepsis and for septic shock.

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
Statistical risk model
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 = Lower 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.)*
Setting

The study objective was to develop a logistic regression model to estimate the probability of hospital mortality among septic patients entering 179 New York State hospitals over the period of January 1, 2015 through December 31, 2015. The a priori analysis plan eliminated any patient with an advanced directive or who declined interventions. When a patient was discharged from a hospital as "transfer to acute care", only the patient's data from the receiving hospital was used in the dataset. If a patient was in the dataset multiple times for sepsis, only the final admission was used. This preserved the outcome of interest (mortality) and observation independence in the data file for developing logistic regression models. This resulted in a database total of 43,204 septic patients. The a priori analysis used only patient demographics, comorbidities, and admission characteristics to estimate the probability of hospital mortality. Specifically treatment variables were not used in the model.

Septic patients

All subjects entered into the model met the admitting hospital's criteria for severe sepsis or septic shock. Severe sepsis was defined as a suspected or confirmed infection, at least two systemic manifestations of infection and one or more acute organ dysfunctions. Septic shock was defined as severe sepsis where at least one organ dysfunction with sustained hypotension after a fluid challenge. For this paper, the term sepsis or septic represents the dataset population of severe sepsis and septic shock patients. Mortality is defined as in-hospitals deaths.

Statistical Methods

Logistic regression developed a model to estimate the probability of mortality for patients with severe sepsis or septic shock during their hospital stay. A list of the possible predictor variables and definitions are given in Table 1. Maximum likelihood was used to estimate model coefficients and associated standard errors. The hierarchical nature of the data supports random-effects logistic regression use since patients are nested within the 179 hospitals. However, the 179 random-effect coefficients would have made the resulting model specific only to those 179 New York hospitals and would not be generalizable to patients outside these specific hospitals. A random sample of 10% (N = 4,319) of the observations were set aside and the logistic regression model was developed on the remaining 90% (38,884 observations). The final model was validated on the 10% of observations that were set aside. Patient comorbidities were generated using the list shown in supplemental Table S1. We generated a variable called mechanical ventilation (MV) severity that indicated a severity of illness relating to mechanical ventilation. This dichotomous variable was defined when a patient was admitted to the hospital already mechanically ventilated or requiring mechanical ventilation within 6 hours post admission. Initial serum lactate was not measured in 2,528 (5.9%) patients and was imputed using single imputation. Specifically, truncated linear regression was used during the imputation procedure where the lower limit of left truncation was set at a serum lactate level 0.1 mmol/L (1st percentile) and the upper limit of the right truncation was set at 30.0 mmol/L (99th percentile). A list of predictor variables is shown in supplemental Table S2.

A multivariable logistic regression model was built using the developmental dataset and starting with all possible covariates in the model. Using an iterative procedure, variables were removed from the model, one by one, if their p-values were not significant at 0.05 level until a parsimonious model was reached. Variables removed during the development procedure were added back into the model if their p-values were significant at the 0.05 level and if model calibration (Hosmer-Lemeshow goodness of fit) was improved through their inclusion. We then assessed the scale of the 3 continuous variables (patient age, first serum lactate, and the count of the number of comorbidities) remaining in the model. Specifically, we were interested in determining whether these variables had a linear relationship with mortality. Using the method of fractional polynomials patient age was included in the model as a linear term, the number of comorbidities was transformed by taking the square root of the number of comorbidities, and first serum lactate was entered into the model as a quadratic expression (linear and a squared term). Model calibration was further improved by adding the following interactions to the model: lower respiratory infection (LRI) and MV severity, patient age and the square root of the number of comorbidities, and first serum lactate and the square root of the number of comorbidities.

Model calibration was assessed using the Hosmer-Lemeshow goodness of fit on both the developmental and the validation datasets. Group sizes of 10, 100, 500, and 1,000 were chosen for the large, developmental, dataset while group sizes of 10, 50, 100, and 150 were chosen for the smaller validation dataset. Model discrimination was assessed using the area under the receiver operating characteristic (ROC) curve for both the developmental and validation datasets.

The estimated probability of mortality was generated using the model coefficients and the specific patient attributes. If the patient attribute is defined by a categorical variable, then the possible values are either a 0 or 1. If the attribute is defined by a continuous variable, then the specific value is used such as the patient's age. Interaction values are generated by multiplying the values of each of the two individual variables defined by the interaction. The product of the coefficient and the patient's value for all of the variables in the model are generated. Next the logit is defined as the sum of the above products. Finally, the probability of mortality for a specific patient is generated using the follow equation:

Probability of mortality= exp(logit)/(1+exp(logit))

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.

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.

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Claims (Other), EHRs Hybrid, Laboratory, Management Data, Non-Medical Data, Paper Records, Pharmacy, Registry

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.
Data collection is performed via a standardized clinical data dictionary (see Appendix) with set specified data fields which may be electronically extracted via custom record abstraction queries and/or manually abstracted, all of which conclude with a plain-text comma-delimited file. The file is submitted over a secure encrypted connection to an electronic data collection portal (https://ny.sepsis.ipro.org) that validates all data and all conditional bounds of data subject to an electronic machine-readable version of the data dictionary which parses not only valid data but also ensures that all "if then" statements are conditionally valid, e.g. ""left_ed_datetime cannot be before triage_datetime"". All required data elements must be completed for the submission to be accepted by the portal. Data errors such as conditional logic failures or missing data are returned to the submitter for correction prior to data acceptance. The portal maintains valid dictionaries for all reporting periods such that historical data may be submitted and validated against historical versions of the data dictionary.

Valid data is passed on to the analytic process, invalid data is destroyed and an error returned to the submitter with detailed failure reasons and a requirement to resubmit the data upon correction. Full data submission is validated through facility volume comparison charts across prior data quarters and years.

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)*
Available in attached appendix at A.1

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)*
Hospital : Acute Care Facility
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.)*

2. Validity – See attached Measure Testing Submission Form
NQF_nqf_testing_attachment_01302017.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.)
Yes

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)
Yes - Updated information required during the SDS Trial Period is included

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

Measure Number (*if previously endorsed*)**:** 3215
Measure Title: Adult Risk Adjusted Sepsis Mortality
Date of Submission: Click here to enter a date
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.

- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing, (e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☒ administrative claims	☐ administrative claims

☒ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Hospital self-report data using standardized data dictionary and file structure. Data could come from EHR, admin claims, paper records, registry, etc.	☒ other: A 10% random sample of cases were externally validated through medical record review to ensure data accuracy.

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*).

New York State began a statewide initiative to accelerate care improvements for severe sepsis and septic shock patients with the goal of reducing variations in the delivery of evidence informed care and preventable mortality across New York hospitals. Hospitals were required to develop and implement recognition and treatment protocols for septic patients at the inception of disease state. Clinical data submitted to the Department of Health (DOH) on a quarterly basis is evaluated for protocol use, adherence to timed interventions, and patient outcomes including mortality. The submitted data is audited for accuracy through random sample chart review. Additionally clinical cases are compared to administrative data submission for all patients and all discharges to determine completeness. From this above dataset we used 43,204 severe sepsis and septic shock patients from 179 hospitals January 1, 2015 to December 31, 2015 to create a model that estimated the probability of hospital mortality. The model was developed using 38,884 (90%) observations and the remaining 4,319 (10%) were used to validate the data.

1.3. What are the dates of the data used in testing? Discharges within calendar year 2015 were included in the measure.

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

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

1.5. How many and which <u>measured entities</u> 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*)

A 90% random sample was drawn from the entire dataset that was used to develop the logistic regression model where the remaining 10% was used as a validation dataset. Specifically we used our statistical software (Stata, version 14.2) that uses the command "*sample*" to takes a random sample without replacement using a random seed. The random seed allows us reproduce the same random sample in the future when needed. The table below give a comparison of the model outcome (hospital mortality) and all of the variables used to predict hospital mortality across the validation and developmental datasets. As expected the percentages in the validation and the developmental datasets are very close to each other.

Table 1: Patient characteristics across the validation and developmental datasets

Patient characteristics	**Validation data N = 4,320 (10%)**		**Development data N = 38,884 (90%)**		**Total dataset N = 43,204 (100%)**	
	N	**Percent**	**N**	**Percent**	**N**	**Percent**
Number of unique hospitals	160		179			179
Hospital mortality	1,313	30.4	11,741	30.2	13,054	30.2
Median age, (IQR)	72	(59 - 83)	71	(59 - 82)	71	(59 - 82)
Race/Ethnicity						
White, non-Hispanic	2,454	56.8	22,041	56.7	24,495	56.7
Black, non-Hispanic	666	15.4	5,938	15.3	6,604	15.3
Hispanic	389	9.0	3,603	9.3	3,992	9.2
Multi-racial	47	1.1	627	1.6	674	1.6
Unknown, non-Hispanic	332	7.7	2,783	7.2	3,115	7.2
Unknown	432	10.0	3,892	10.0	4,324	10.0
Payer						
Medicare	2,585	59.8	23,308	59.9	25,893	59.9
Medicaid	737	17.1	6,455	16.6	7,192	16.7
Private, HMO	816	18.9	7,363	18.9	8,179	18.9
Self-Pay	50	1.2	503	1.3	553	1.3
Other	132	3.1	1,255	3.2	1,387	3.2
Admission source						
Non-health facility	2,918	67.6	26,557	68.3	29,475	68.2
Clinic	174	4.0	1,610	4.1	1,784	4.1
Different Hospital	268	6.2	2,445	6.3	2,713	6.3
SNF/ICF	886	20.5	7,577	19.5	8,463	19.6
Another health care facility	34	0.8	325	0.8	359	0.8
Between unit transfer	22	0.5	162	0.4	184	0.4
Hospice	1	0.0	28	0.1	29	0.1
Other	17	0.4	180	0.5	197	0.5
Lower respiratory infection	2,097	48.5	19,346	49.8	21,443	49.6
MV Severity	516	11.9	4,841	12.5	5,357	12.4
Severe Sepsis	2,204	51.0	19,939	51.3	22,143	51.3
Septic Shock	2,116	49.0	18,945	48.7	21,061	48.8
Thrombocytopenia	1,071	24.8	9,155	23.5	10,226	23.7
Metastatic cancer	442	10.23	4,162	10.7	4,604	10.7

Patient characteristics	Validation data N = 4,320 (10%)		Development data N = 38,884 (90%)		Total dataset N = 43,204 (100%)	
	N	Percent	N	Percent	N	Percent
Lymphoma/Leukemia/Multiple Myeloma	218	5.1	2,018	5.2	2,236	5.2
Median serum lactate, (IQR)	2.5	(1.5-4.3)	2.5	(1.5-4.2)	2.5	(1.5-4.2)
Median number of comorbidities, (IQR)	3	(1 - 4)	3	(1 - 4)	3	(1 - 4)

1.6. How many and which <u>patients</u> 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*)

Please see Table 1 in section 1.5

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.

There were no notable differences between the validation sample and the full population used for generation of the model for risk adjustment.

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).

Patient captured demographic variables were all included for model building through not all demographic data elements remained viable in the final model. The demographic variables under consideration included: age; gender; payor; race; and, ethnicity. Inclusion in the model was based on whether or not the demographic variable was statistically significant at the 0.05 level. Gender was the only variable not to be included in the model since its odds ratio and corresponding p-value was 1.0003 and 0.992, respectively. All of the other demographic variables had p-values < 0.001 for at least one of the levels of a specific demographic (see table in section 2b4.1.1)

2a2. RELIABILITY TESTING

***<u>Note</u>**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

There are two components of reliability testing for these data. One component compares the accuracy of the data submission from the hospitals to independent external medical record abstraction. These results are presented in the table as "Audit Results" in column one. The other component compares the performance of the auditors in the ability to validly and reliably abastract the hospital data from the medical record. Gold standard testing required 100% accuracy prior to initiating reliability testing. These results are presented in the table below in columns two through four.

Variable Name	Audit Results (% Validity)	IRR Gold Standard	IRR POA	IRR Kappa
Race	N/A	N/A	N/A	N/A
Ethnicity	N/A	N/A	N/A	N/A
Payer	N/A	N/A	N/A	N/A
Site of Infection	98.9%	100.0%	100.00	SC
Admission Source	N/A	100.0%	93.33	0.6341
Lower Respiratory Infection	98.8%	100.0%	93.33	SC
Mechanical Ventilation	97.8%	100.0%	100	SC
Age (Date of Birth)	99.2%	100.0%	96.67	SC
Thrombocytopenia	97.7%	100.0%	93.33	SC
Septic Shock	98.4%	100.0%	100.00	SC
Serum Lactate (Lactate Level)	93.9%	100.0%	96.67	0.8519
Metastatic Cancer	97.1%	100.0%	100.00	1
Lymphoma, Leukemia, Multiple Myeloma	99.1%	100.0%	96.67	SC
Square Root of Comorbidity Count (Range of Comorbidities)	89.9%-99.1%	100.0%	96.67-100.00	SC
*N/A=These variables were not audited manually but were aligned to state administrative datasets to ensure accuracy. SC=There were insufficient cell spread to enable kappa calculation. This can happen with high percent agreement with the same response category.				

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*)

Hospitals submitted hospital-abstracted data. The Department audited the data prior to the creation of process or outcome measures (results as noted above). The Department reviewers were trained to follow the data dictionary according to definitions provided within the document. No reviewer (data auditor) was authorized to begin review outside of training cases without having met IRR results exceeding Fleiss and Cohen standards with > 90% agreement with kappa >.60.

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

Industry standards were met with the above demonstrated results ensuring that the hospital self-reported data was within acceptable accuracy standards and can be interpretated as reliable and accurate data. When

adjusting for chance, agreement percentages and kappa scores were above statistical standards for performance.

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)*

This is the first sepsis-specific risk model and therefore unique in class. The only other existing model built for sepsis (Osborn TM et al. Crit Care Med 2014;42(9):1969-76) included multiple treatment variables, which prevented widespread application of the model to other databases. In addition, there are no other large sepsis databases on which to test this current risk model, apart from the CMS sepsis database, which is not yet available.

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

Empirical validity testing:

- The Hosmer-Lemeshow goodness of fit test for the developmental and the validation datasets indicate:
  - For the developmental dataset, we used group sizes of 10, 100, 500, and 1,000 and the *p*-values for the tests were 0.568, 0.972, 0.735, and 0.735, respectively.
  - For the validation dataset we used group sizes of 10, 50, 100, and 150 and the *p*-values for the tests were 0.651, 0.977, 0.985, and 0.974, respectively.
- The ROC area for the developmental and validation datasets are 0.770 and 0.773, respectively, indicating good discrimination

Systematic face validity:

The adjusted probability of hospital mortality are shown in Table 1 below along with the adjusted hospital mortality odds ratios.

Table 1

Main effects or Interactions	Adjusted mortality, %	Adjusted OR	OR: 95% CI		*p*-value
Race/Ethnicity					
White, non-Hispanic (Referent)	29.5	1.00			
Black, non-Hispanic	32.8	1.21	1.12	1.30	< 0.001
Hispanic	28.3	0.93	0.85	1.02	0.112
Multi-racial	32.4	1.18	0.97	1.43	0.096
Unknown, non-Hispanic	30.3	1.04	0.95	1.15	0.382
Unknown	31.3	1.11	1.02	1.20	0.016
Payer					

Main effects or Interactions	Adjusted mortality, %	Adjusted OR	OR: 95% CI		*p*-value
Medicare (Referent)	29.6	1.00			
Medicaid	31.3	1.11	1.02	1.19	0.011
Private HMO	30.8	1.07	1.00	1.15	0.056
Self-pay	41.3	1.91	1.53	2.39	< 0.001
Other	29.2	0.98	0.85	1.13	0.750
Site of infection					
Urinary (Referent)	22.2	1.00			
Respiratory	32.0	1.82	1.68	1.97	< 0.001
Gastrointestinal	31.0	1.72	1.57	1.88	< 0.001
Skin	30.0	1.62	1.44	1.82	< 0.001
Central nervous system	34.2	2.06	1.49	2.86	< 0.001
Other	31.3	1.75	1.58	1.93	< 0.001
Unknown	37.5	2.46	2.23	2.71	< 0.001
Admission source					
Non-health facility, POA (Referent)	28.6	1.00			
Clinic	28.1	0.97	0.86	1.10	0.605
Different Hospital	36.0	1.52	1.38	1.67	< 0.001
Skilled nursing facility/Intermediate care facility	33.6	1.34	1.26	1.42	< 0.001
Another health care facility	33.4	1.32	1.01	1.72	0.040
Between unit transfer	36.7	1.58	1.10	2.28	0.013
Hospice	41.8	2.07	0.93	4.61	0.075
Other	23.4	0.72	0.48	1.07	0.106
Septic shock vs. severe sepsis	36.2 vs. 22.7	2.16	2.04	2.29	< 0.001
Thrombocytopenia vs. no thrombocytopenia	33.9 vs. 29.0	1.33	1.26	1.41	< 0.001
Metastatic cancer vs. none	37.4 vs. 29.2	1.58	1.47	1.71	< 0.001
Lymphoma/leukemia/multiple myeloma vs. none	32.4 vs. 30.1	1.14	1.03	1.27	0.011
MV severity significantly interacts with lower respiratory infection thus there are four possible odds ratios					
Lower respiratory infection vs. none and the patient has no MV severity	31.8 vs. 27.0	1.33	1.24	1.41	< 0.001
Lower respiratory infection vs. none and the patient has MV severity	33.9 vs. 36.1	0.89	0.78	1.02	0.089
MV severity vs. no MV severity and the patient has no lower respiratory infection	36.1 vs. 27.0	1.68	1.49	1.89	< 0.001
MV severity vs. no MV severity and the patient has lower respiratory infection	33.9 vs. 31.8	1.13	1.03	1.23	0.007
Ten year increase in age at the following number of comorbidities since age significantly interacts with square root of the number of comorbidities					
80 vs. 70 years of age with 0 comorbidities	18.9 vs. 12.3	1.81	1.71	1.92	< 0.001
80 vs. 70 years of age with 2 comorbidities	31.4 vs. 26.1	1.35	1.32	1.37	< 0.001
80 vs. 70 years of age with 4 comorbidities	38.0 vs. 34.5	1.19	1.16	1.21	< 0.001
80 vs. 70 years of age with 6 comorbidities	43.5 vs. 41.8	1.08	1.05	1.12	< 0.001
One unit increase in the number of comorbidities for specific combinations of patient age and their serum lactate value since the square root of the number of comorbidities significantly interacts with age and significantly interacts with serum lactate					
1 vs. 0 comorbidities at age = 50, and serum lactate = 2 mmol/L	8.6 vs. 2.7	3.56	3.20	3.95	< 0.001
1 vs. 0 comorbidities at age = 50, and serum lactate = 3 mmol/L	10.2 vs. 3.3	3.41	3.07	3.78	< 0.001
1 vs. 0 comorbidities at age = 50, and serum lactate = 4 mmol/L	12.0 vs. 4.1	3.27	2.95	3.62	< 0.001

Main effects or Interactions	Adjusted mortality, %	Adjusted OR	OR: 95% CI		*p*-value
1 vs. 0 comorbidities at age = 60, and serum lactate = 2 mmol/L	12.0 vs. 4.7	2.88	2.65	3.13	< 0.001
1 vs. 0 comorbidities at age = 60, and serum lactate = 3 mmol/L	14.1 vs. 5.8	2.76	2.55	2.99	< 0.001
1 vs. 0 comorbidities at age = 60, and serum lactate = 4 mmol/L	16.4 vs. 7.2	2.64	2.44	2.86	< 0.001
1 vs. 0 comorbidities at age = 70, and serum lactate = 2 mmol/L	16.5 vs. 8.1	2.33	2.18	2.49	< 0.001
1 vs. 0 comorbidities at age = 70, and serum lactate = 3 mmol/L	19.1 vs. 9.9	2.23	2.10	2.38	< 0.001
1 vs. 0 comorbidities at age = 70, and serum lactate = 4 mmol/L	22.0 vs. 12.0	2.14	2.01	2.28	< 0.001
1 vs. 0 comorbidities at age = 80, and serum lactate = 2 mmol/L	22.1 vs. 13.5	1.89	1.77	2.02	< 0.001
1 vs. 0 comorbidities at age = 80, and serum lactate = 3 mmol/L	25.3 vs. 16.3	1.81	1.70	1.92	< 0.001
1 vs. 0 comorbidities at age = 80, and serum lactate = 4 mmol/L	28.7 vs. 19.4	1.73	1.63	1.84	< 0.001
2 vs. 1 comorbidities at age = 50, and serum lactate = 2 mmol/L	13.6 vs. 8.6	1.69	1.62	1.77	< 0.001
2 vs. 1 comorbidities at age = 50, and serum lactate = 3 mmol/L	15.6 vs. 10.2	1.66	1.59	1.73	< 0.001
2 vs. 1 comorbidities at age = 50, and serum lactate = 4 mmol/L	17.8 vs. 12.0	1.63	1.56	1.70	< 0.001
2 vs. 1 comorbidities at age = 60, and serum lactate = 2 mmol/L	17.2 vs. 12.0	1.55	1.50	1.60	< 0.001
2 vs. 1 comorbidities at age = 60, and serum lactate = 3 mmol/L	19.7 vs. 14.1	1.52	1.47	1.57	< 0.001
2 vs. 1 comorbidities at age = 60, and serum lactate = 4 mmol/L	22.3 vs. 16.4	1.50	1.45	1.55	< 0.001
2 vs. 1 comorbidities at age = 70, and serum lactate = 2 mmol/L	21.6 vs. 16.5	1.42	1.38	1.46	< 0.001
2 vs. 1 comorbidities at age = 70, and serum lactate = 3 mmol/L	24.4 vs. 19.1	1.40	1.36	1.43	< 0.001
2 vs. 1 comorbidities at age = 70, and serum lactate = 4 mmol/L	27.4 vs. 22.0	1.37	1.34	1.41	< 0.001
2 vs. 1 comorbidities at age = 80, and serum lactate = 2 mmol/L	26.6 vs. 22.1	1.30	1.27	1.34	< 0.001
2 vs. 1 comorbidities at age = 80, and serum lactate = 3 mmol/L	29.8 vs. 25.3	1.28	1.25	1.31	< 0.001
2 vs. 1 comorbidities at age = 80, and serum lactate = 4 mmol/L	33.2 vs. 28.7	1.26	1.22	1.29	< 0.001
3 vs. 2 comorbidities at age = 50, and serum lactate = 2 mmol/L	18.7 vs. 13.6	1.50	1.45	1.55	< 0.001
3 vs. 2 comorbidities at age = 50, and serum lactate = 3 mmol/L	21.1 vs. 15.6	1.48	1.43	1.53	< 0.001
3 vs. 2 comorbidities at age = 50, and serum lactate = 4 mmol/L	23.6 vs. 17.8	1.46	1.41	1.51	< 0.001
3 vs. 2 comorbidities at age = 60, and serum lactate = 2 mmol/L	22.2 vs. 17.2	1.40	1.36	1.44	< 0.001
3 vs. 2 comorbidities at age = 60, and serum lactate = 3 mmol/L	24.9 vs. 19.7	1.38	1.35	1.42	< 0.001
3 vs. 2 comorbidities at age = 60, and serum lactate = 4 mmol/L	27.6 vs. 22.3	1.36	1.33	1.40	< 0.001
3 vs. 2 comorbidities at age = 70, and serum lactate = 2 mmol/L	26.1 vs. 21.6	1.31	1.28	1.34	< 0.001
3 vs. 2 comorbidities at age = 70, and serum lactate = 3 mmol/L	29.1 vs. 24.4	1.29	1.27	1.32	< 0.001

Main effects or Interactions	Adjusted mortality, %	Adjusted OR	OR: 95% CI		*p*-value
3 vs. 2 comorbidities at age = 70, and serum lactate = 4 mmol/L	32.1 vs. 27.4	1.27	1.25	1.30	< 0.001
3 vs. 2 comorbidities at age = 80, and serum lactate = 2 mmol/L	30.4 vs. 26.6	1.22	1.20	1.25	< 0.001
3 vs. 2 comorbidities at age = 80, and serum lactate = 3 mmol/L	33.6 vs. 29.8	1.21	1.18	1.23	< 0.001
3 vs. 2 comorbidities at age = 80, and serum lactate = 4 mmol/L	36.8 vs. 33.2	1.19	1.17	1.21	< 0.001
4 vs. 3 comorbidities at age = 50, and serum lactate = 2 mmol/L	24.0 vs. 18.7	1.40	1.37	1.45	< 0.001
4 vs. 3 comorbidities at age = 50, and serum lactate = 3 mmol/L	26.6 vs. 21.1	1.39	1.35	1.43	< 0.001
4 vs. 3 comorbidities at age = 50, and serum lactate = 4 mmol/L	29.3 vs. 23.6	1.37	1.34	1.41	< 0.001
4 vs. 3 comorbidities at age = 60, and serum lactate = 2 mmol/L	27.1 vs. 22.2	1.33	1.30	1.36	< 0.001
4 vs. 3 comorbidities at age = 60, and serum lactate = 3 mmol/L	29.9 vs. 24.9	1.31	1.28	1.34	< 0.001
4 vs. 3 comorbidities at age = 60, and serum lactate = 4 mmol/L	32.7 vs. 27.6	1.30	1.27	1.33	< 0.001
4 vs. 3 comorbidities at age = 70, and serum lactate = 2 mmol/L	30.4 vs. 26.1	1.25	1.23	1.28	< 0.001
4 vs. 3 comorbidities at age = 70, and serum lactate = 3 mmol/L	33.3 vs. 29.1	1.24	1.22	1.26	< 0.001
4 vs. 3 comorbidities at age = 70, and serum lactate = 4 mmol/L	36.3 vs. 32.1	1.23	1.21	1.25	< 0.001
4 vs. 3 comorbidities at age = 80, and serum lactate = 2 mmol/L	33.8 vs. 30.4	1.19	1.16	1.21	< 0.001
4 vs. 3 comorbidities at age = 80, and serum lactate = 3 mmol/L	36.9 vs. 33.6	1.17	1.15	1.19	< 0.001
4 vs. 3 comorbidities at age = 80, and serum lactate = 4 mmol/L	40.0 vs. 36.8	1.16	1.14	1.18	< 0.001
One unit increase in serum lactate at the following comorbidities since the square root of the number of comorbidities significantly interacts with serum lactate					
2 vs. 1 mmol/L serum lactate with 0 comorbidities	17.4 vs. 8.1	1.27	1.23	1.31	< 0.001
3 vs. 2 mmol/L serum lactate with 0 comorbidities	24.9 vs. 17.4	1.26	1.22	1.30	< 0.001
4 vs. 3 mmol/L serum lactate with 0 comorbidities	32.2 vs. 24.9	1.26	1.22	1.30	< 0.001
2 vs. 1 mmol/L serum lactate with 1 comorbidities	20.1 vs. 9.9	1.22	1.19	1.24	< 0.001
3 vs. 2 mmol/L serum lactate with 1 comorbidities	27.8 vs. 20.1	1.21	1.19	1.23	< 0.001
4 vs. 3 mmol/L serum lactate with 1 comorbidities	27.5 vs. 27.8	1.20	1.18	1.23	< 0.001
2 vs. 1 mmol/L serum lactate with 2 comorbidities	22.9 vs. 12.0	1.19	1.17	1.22	< 0.001
3 vs. 2 mmol/L serum lactate with 2 comorbidities	23.5 vs. 22.9	1.19	1.17	1.21	< 0.001
4 vs. 3 mmol/L serum lactate with 2 comorbidities	30.4 vs. 23.5	1.18	1.17	1.20	< 0.001
2 vs. 1 mmol/L serum lactate with 3 comorbidities	19.4 vs. 14.3	1.18	1.16	1.20	< 0.001
3 vs. 2 mmol/L serum lactate with 3 comorbidities	26.3 vs. 19.4	1.17	1.16	1.19	< 0.001

Main effects or Interactions	Adjusted mortality, %	Adjusted OR	OR: 95% CI		*p*-value
4 vs. 3 mmol/L serum lactate with 3 comorbidities	33.3 vs. 26.3	1.17	1.15	1.18	< 0.001
2 vs. 1 mmol/L serum lactate with 4 comorbidities	22.0 vs. 15.0	1.16	1.14	1.19	< 0.001
3 vs. 2 mmol/L serum lactate with 4 comorbidities	29.2 vs. 22.0	1.16	1.14	1.18	< 0.001
4 vs. 3 mmol/L serum lactate with 4 comorbidities	36.3 vs. 29.2	1.15	1.14	1.17	< 0.001

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?*)

Empirical validity testing:

- The Hosmer-Lemeshow goodness of fit test *p*-values for both the developmental and the validation datasets are all > 0.05 indicating that there is no evidence of lack of fit. The the null hypothesis for the Hosmer-Lemeshow goodness of fit test is that the model estimates of the probability of mortality match the observed/actual mortality. Thus, a non-significant *p*-value fit.
- The area under the ROC curve is greater than 0.7 for both the developmental and the validation datasets indicating good model discrimination.

Systematic face validity:

- All of the probabilities of hospital mortality are higher in the patients that have a particular comorbidity, higher in those that are older, higher in those whose first serum lactate is higher, and higher in those with more comorbidities. Additionally all of the odds ratios are higher in these same groups than the lower risk groups.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section ___*

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*)

We excluded a patient's prior addmissions and kept their final hospital admission for those patients with multiple addmissions to New York State hospitals during the study period. This was done for two statistical reasons. First we needed the outcome to be statistically independent that is a necessary assumption when running a logistic regression model. Secondly if a patient has multiple hospital admissions for sepsis, it is impossible for them to have had the outcome interst (hospital mortality) in any but their last admission. We also excluded patients that had an advanced directives in place that precluded one or more elements of the sepsis protocol. Additionally we excluded if the patient or surrogate decision maker declined interventions.

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*)

We did not test the exclusions for the following reasons:

- Keeping prior hospital admissions in the dataset violates the principle of statistical independence of the outcome a necessary assumption when using maximum likelihood logistic regression
- Keeping patients in the dataset that had an advanced directive or declined intervention would not be appropriate since the population of interest are those that were eligible to receive the treatment.

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*)

We feel that keeping those excluded patients in the analysis would bias results as follows:

- Keeping prior hospital admissions in the dataset would bias the results towards lower hospital mortality since it is impossible to experience hospital mortality since they obviously survived in order to be admitted to hospital (i.e. their last observation).
- Keeping patients in the dataset that had an advanced directive or declined intervention would bias the results towards higher hospital mortality.

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 ____.

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

☐ **No risk adjustment or stratification**

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

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

☐ **Other,** Click here to enter description

2b4.1.1 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.

The risk adjusted hospital mortality logistic regressionmodel is shown below in Table 1. The terms/variables in the model are described in the data dictionary

Table 1: Logistic regression model for hospital mortality

Term in the logistic regression model	β	β: 95% CI		SE	*p*-value
Race/ethnicity					
White, non-Hispanic (Referent)	0.000				
Black, non-Hispanic	0.188	0.116	0.259	0.036	< 0.001
Hispanic	-0.073	-0.163	0.017	0.046	0.112
Multi-racial	0.165	-0.029	0.359	0.099	0.096
Unknown, non-Hispanic	0.043	-0.053	0.139	0.049	0.382
Unknown	0.101	0.019	0.183	0.042	0.016
Payer					
Medicare (Referent)	0.000				
Medicaid	0.099	0.023	0.176	0.039	0.011
Private, HMO	0.067	-0.002	0.136	0.035	0.056
Self-Pay	0.646	0.422	0.870	0.114	< 0.001
Other	-0.023	-0.167	0.120	0.073	0.750

Term in the logistic regression model	β	β: 95% CI		SE	*p*-value
Site of infection					
Urinary (Referent)	0.000				
Respiratory	0.599	0.521	0.677	0.040	< 0.001
Gastrointestinal	0.543	0.454	0.632	0.045	< 0.001
Skin	0.484	0.367	0.600	0.059	< 0.001
Central Nervous System	0.723	0.396	1.050	0.167	< 0.001
Other	0.558	0.457	0.659	0.052	< 0.001
Unknown	0.900	0.802	0.997	0.050	< 0.001
Admission source					
Non-health facility, POA (Referent)	0.000				
Clinic	-0.033	-0.157	0.091	0.063	0.605
Different Hospital	0.418	0.322	0.515	0.049	< 0.001
Skilled nursing facility/Intermediate care facility	0.289	0.227	0.351	0.032	< 0.001
Another health care facility	0.278	0.012	0.544	0.136	0.040
Between unit transfer	0.459	0.096	0.823	0.186	0.013
Hospice	0.727	-0.073	1.528	0.408	0.075
Other	-0.330	-0.730	0.070	0.204	0.106
Lower respiratory infection					
No (Referent)	0.000				
Yes	0.281	0.217	0.345	0.033	< 0.001
MV severity					
No (Referent)	0.000				
Yes	0.519	0.402	0.636	0.060	< 0.001
Lower respiratory infection and MV severity	-0.398	-0.538	-0.258	0.072	< 0.001
Septic shock diagnosis					
Severe Sepsis	0.000				
Septic Shock	0.770	0.712	0.828	0.029	< 0.001
Thrombocytopenia					
No (Referent)	0.000				
Yes	0.285	0.229	0.341	0.029	< 0.001
Metastatic cancer					
No (Referent)	0.000				
Yes	0.460	0.385	0.535	0.038	< 0.001
Lymphoma/leukemia/multiple myeloma					
No (Referent)	0.000				
Yes	0.135	0.030	0.239	0.053	0.011
Patient age	0.060	0.054	0.065	0.003	< 0.001
Square root of comorbidity count	2.410	2.153	2.667	0.131	< 0.001
Age and the square root of comorbidity count	-0.021	-0.024	-0.018	0.002	< 0.001
First serum lactate	0.245	0.210	0.279	0.018	< 0.001
First serum lactate squared	-0.002	-0.004	-0.001	0.001	0.001
Serum lactate and the square root of comorbidity count	-0.043	-0.059	-0.026	0.008	< 0.001
Constant term	-8.548	-9.018	-8.078	0.240	< 0.001

Empirical validity testing:

The goal of the risk adjusted logistic regression model was to produce probabilities of hospital mortality that accurately reflect the actual mortality experience of the septic patients. Three techniques were used to generate a valid risk adjusted hospital mortatily model

1. A random sample of 10% (N = 4,319) of the observations were set aside and the logistic regression model was developed on the remaining 90% (38,884 observations). The final model was validated on the 10% of observations that were set aside.
2. The Hosmer-Lemeshow goodness of fit test was used to assess model calibration. This test compares the number of deaths observed as compared to the number of deaths estimsted (basedon the model) within a fixed number of subgroups (based on the estimated probabilities). For example, use of 10 groups would create deciles of estimated probabilities starting with the minimum and ending with the maximum probability. Model calibration was assessed on both the developmental and the validation datasets. Group sizes of 10, 100, 500, and 1,000 were chosen for the large, developmental, dataset while group sizes of 10, 50, 100, and 150 were chosen for the smaller validation dataset.
3. Model discrimination was assessed using the area under the receiver operating characteristic (ROC) curve for both the developmental and validation datasets. The closer the area under the ROC curve is to 1.0 the better the model can discriminate between those patients that die (by estimating high probabilities of dying) during there hospital stay and those that do not die (by estimating low probabilities of dying).

Systematic face validity:

Two methods were used to assess the face validity of the risk adjusted logistic regression model.

a. The probability of hospital mortality was generated for each covariate in the model. For example, the variable septic shock diagnosis was used as a covariate in the model and face validity was checked by determining whether the probability of mortality for those patients without septic shock was higher than the probability of mortality for severe sepsis. These probabilities are adjusted for the other variables in the model. Specifically the probabilities are generated after running the logistic regression model for each level of the of the covariate of interest (e.g. septic shock diagnosis) while integrating (averaging) over the remaining covariates.
b. The face validity of the risk adjusted mortality model was assessed by assuring that the adjusted hospital mortality odds ratio was greater than 1.0 for a particular covariate. For example the odds ratio for septic shock diagnosis should be greater than 1.0 as a measure of face validity. That is, the odds of hospital mortality for a diagnosis of septic shock should be higher than the odds of hospital mortality for a diagnosis of severe sepsis. After running the logistic regression model, odds ratios are produced by specifying specific linear contrast statements using the variables of interest.

2b4.2. If an outcome or resource use component measure is <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical <u>and</u> 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*)

A multivariable logistic regression model was built using the developmental dataset and starting with all possible covariates in the model. Using an iterative procedure, variables were removed from the model, one by one, if their *p*-values were not significant at 0.05 level until a parsimonious model was reached. Variables removed during the development procedure were added back into the reduced model if their *p*-values were significant at the 0.05 level and if model calibration (Hosmer-Lemeshow goodness of fit) was improved through their inclusion. We then assessed the scale of the 3 continuous variables (patient age, first serum lactate, and the count of the number of comorbidities) remaining in the model. Specifically, we were interested in

determining whether these variables had a linear relationship with mortality (i.e., linear in the logit). Using the method of fractional polynomials patient age was included in the model as a linear term, number of comorbidities was transformed by taking the square root, and first serum lactate was entered into the model as a quadratic expression (linear and a squared term). Model calibration was further improved by adding the following interactions to the model: lower respiratory infection (LRI) by MV severity, patient age by the square root of the number of comorbidities, and first serum lactate by the square root of the number of comorbidities.

2b4.4a. What were the statistical results of the analyses used to select risk factors?
See response to 2b4.3

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)
There were no Sociodemographic Status (SDS) variables collected by New York State during this study. Variables reflecting patient demographics were collected (e.g., age, gender, race, and ethnicity). To be included in the model these demographic variables followed the same procedure as outlined in 2b4.3

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*)

1. The Hosmer-Lemeshow goodness of fit test was used to assess model calibration. This is a measure of how closely the estimated probabilities of hospital mortality produced from the model reflect the actual, observed mortality across a specified number of groups. For example, a group size of 10 would create deciles of probability starting with the minimum and ending with the maximum probability. Model calibration was assessed on both the developmental and the validation datasets. Group sizes of 10, 100, 500, and 1,000 were chosen for the large, developmental, dataset while group sizes of 10, 50, 100, and 150 were chosen for the smaller validation dataset.
2. Model discrimination was assessed using the area under the receiver operating characteristic (ROC) curve for both the developmental and validation datasets. The closer the area under the ROC curve is to 1.0 the better the model can discriminate between those patients that die during their hospital stay and those that do not die.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to _____

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

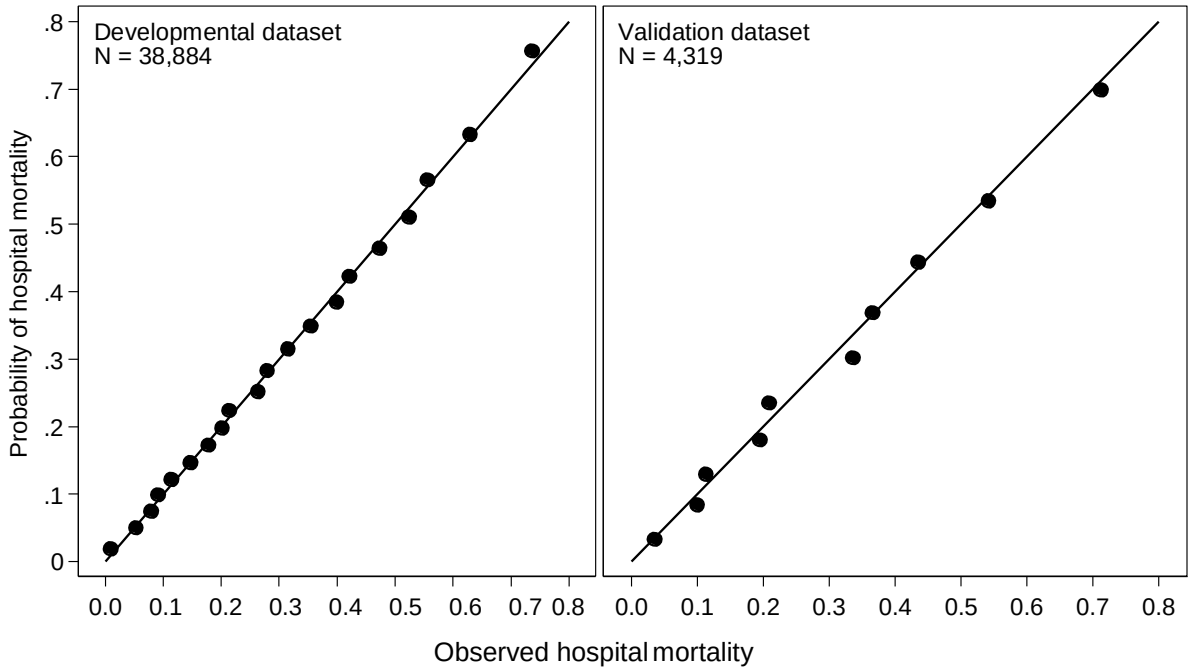
The ROC area for the developmental and validation datasets are 0.770 and 0.773, respectively, indicating good discrimination

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

Hosmer-Lemeshow goodness of fit test for the developmental and the validation datasets indicate:

- For the developmental dataset we used group sizes of 10, 100, 500, and 1,000 and the *p*-values for the tests were 0.568, 0.972, 0.735, and 0.735, respectively.
- For the validation dataset we used group sizes of 10, 50, 100, and 150 and the *p*-values for the tests were 0.651, 0.977, 0.985, and 0.974, respectively.

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



2b4.9. Results of Risk Stratification Analysis:
A risk stratification analysis was not performed

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*)

- The Hosmer-Lemeshow goodness of fit test *p*-values for both the developmental and the validation datasets are all > 0.05 indicating that there is no evidence of lack of fit. The the null hypothesis for the Hosmer-Lemeshow goodness of fit test is that the model estimates of the probability of mortality match the observed/actual mortality. Thus a non-significant *p*-value indicates good fit.
- The area under the ROC curve is greater than 0.7 for both the developmental and the validation datasets indicating good model discrimination.

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

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

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

The primary differences of interest was where or not the logistic regression coefficients were statistically significant. Significance was set at the 0.5 level. Thus coefficients with p-values ≤ 0.05 were considered statistically different from zero. Additionally coefficient p-values were based on the Wald test while comparison of one model vs. another during the development process were based log-likelihood test.

For the Hosmer-Lemeshow goodness of fit test the null hypothesis is that the model estimates of probability hospital mortality are close to observed mortality. Thus p-values that are ≤ 0.05 indicate that the model estimates of hospital mortality are significantly different than the observed mortality. If Hosmer-Lemeshow p-value is > 0.05, then there is no evidence to reject the null hypothesis. Thus the interpretation of a non-significant p-values is that there is no evidence to suggest the model does not fit the data.

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*)

These statistical significant comparisons are shown in section 2.b4

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?*)
Not applicable

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*)

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

2b6.3. What is your interpretation of the results in terms of 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*)

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*)

The patient's first serum lactate was the only variable used in the model with missing data where 6.0% were missing in those that survived their hospital stay and 5.6% were missing in those that died in the hospital. Single imputation using truncated linear regression was used during the imputation procedure where the lower limit of left truncation was set at a serum lactate level 0.1 mmol/L (1st percentile) and the upper limit of the right truncation was set at 30.0 mmol/L (99th percentile). Below are the variables used to impute the missing lactate. Imputing the missing lactate is the preferred statistical approach to handling small amounts (< 10%) of missing data.

Table of predictor variables used to impute the missing serum lactate

Patient characteristics – predictor variables
1. Age
2. Gender
3. Race
4. Ethnicity
5. Payer
6. Admission source
7. Hospital mortality
8. Need for fluids
9. Hypotension
10. Need for vasopressors
11. Thrombocytopenia
12. Bandemia
13. Septic shock vs. severe sepsis
14. Site of infection
15. Mechanical ventilation severity
16. ICU upon hospital admission
17. Chronic respiratory failure
18. Metastatic cancer
19. Lymphoma/Leukemia/Multiple Myeloma
20. Congestive heart failure

21. Chronic renal failure
22. Chronic liver disease
23. Any diabetes
24. Non-missing serum lactate

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*)

Missing lactate	Died		
	No	Yes	Total
No, N	28,351	12,325	40,676
%	94.03	94.42	94.15
Yes, N	1,799	729	2,528
%	5.97	5.58	5.85
Total, N	30,150	13,054	43,204
%	100.00	100.00	100.00

The chi-square *p*-value = 0.120 indicating that we could not find a difference in the percent missing between those that lived and those that died.

When the risk adjusted model was run using the imputed serum lactate the overall risk-adjusted hospital mortality was 29.1%. When the observations associated with missing lactate were dropped from the model development the overall risk adjusted hospital mortality was 29.4%.

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*).

At best, when missing, serum lactate is missing completely at random (MCAR). That is, the missingness is independennt of any other variables in the model. If this were the case then the results would not be biased when the observations associated with the missing serum lactate are dropped. At worst, the missing serum lactate is missing at random (MAR), but where the missingness can be fully accounted for by variables where there is complete information. Thus well known statistical methods can be used to impute the missing serum lactate.

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), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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**.

Some data elements are in defined fields in electronic sources

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 all variables are currently identified or uniformly defined in hospital electronic sources. Many demographic variables are obtainable for hospital electronic health records. These data may be extracted electronically and used in a standard format. Several variables are collected manually by hospitals though some hospitals have created electronic data capture avenues. Often these options differ based on hospital size and resources. NYS has a large volume of acute care hospitals that span, urban/rural, teaching/non-teaching, large/small bed size and the like. It is more likely that hospitals belonging to a larger system of hospitals have been able to general more electronic data capture opportunities. All of the data is submitted following standardized data element definitions, file structure, type and size requirements with algorithms to ensure data complies with known logic criteria as defined previously.

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 collection from hospitals requires systems in place to accurately identify all patients with a diagnosis of severe sepsis or septic shock at some point in their hospital stay. Completion of this task generally requires more than use of administrative or billing codes given known limitations of administrative data with respect to diagnostic accuracy. Further, many data elements will require review and abstraction by trained clinicians and include review of both electronic and paper records at many institutions. Data must be collected by a trusted entity for measurement across institutions with robust mechanisms for ensuring data accuracy and completeness. While sampling can be used for high volume institutions, a random and representative sample must be assured.

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)*.

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 New York State Sepsis Initiative www.health.ny.gov Regulatory and Accreditation Programs New York State Hospital Regulations www.health.ny.gov Quality Improvement (Internal to the specific organization) New York State Department of Health www.health.ny.gov

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

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

Public Reporting
Program/Sponsor: New York State Sepsis Improvement Initiative; New York State Department of Health
Purpose: Transparency and accountability with respect to outcomes of treatment for patients with severe sepsis/septic shock in acute care facilities in NY; quality improvement and recognition of promising practices in care delivery
Geography/Entities/Patients: Statewide; ~180 facilities; ~50,000 patients/year
Level of Measurement/Setting: population based risk adjusted inpatient sepsis mortality within acute care settings

Quality Improvement

Program/Sponsor: New York State Sepsis Improvement Initiative; New York State Department of Health
Purpose: Quality improvement to assist in the recognition of both positive and negative outlier performance, identification of promising practices, analysis of interventions (protocols, bundles, etc.) on patients including subpopulations; used in conjunction with Partnership for Patients care improvement with New York hospitals and New York HENs (hospital associations)
Geography/Entities/Patients: Statewide; ~180 facilities; ~50,000 patients/year
Level of Measurement/Setting: risk adjusted inpatient sepsis mortality within acute care settings

Regulatory Programs

Program/Sponsor: New York State Sepsis Improvement Initiative; New York State Department of Health
Purpose: fulfills Title 10 Regulation (Sections 405.2 and 405.4) in New York State and Section 2805-m of Public Health Law requiring the collection and reporting by the Department of data regarding the performance of hospitals for patients with sepsis including risk adjusted mortality rates for individual hospitals
Geography/Entities/Patients: Statewide; ~180 institutions; ~50,000 patients
Level of Measurement/Setting: risk adjusted inpatient sepsis mortality rates within acute care settings

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?*)

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.*)

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.
Preliminary evidence shows both raw and risk adjusted mortality improvements over time, using data from the second quarter of 2014 to the third quarter of 2016. Public release of New York´s first hospital specific risk adjusted sepsis mortality rates will occur in early 2017 and it will represent hospital performance during calendar year 2015. It is too early to project whether or how this will impact this trend.

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.**

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

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.
The New York Sepsis Advisory Group, an ad hoc open invitation forum for hospitals, clinicians, and associations to provide input to the Department regarding the sepsis initiative, provided input into the development of variables in the data dictionary for

purposes of eventual risk adjustment. This group had (and continues to have) on-going input into the data dictionary iterations for this, and other data purposes. In addition, the Advisory Group reviewed an initial risk adjustment model and were able to make comments and suggestions which were incorporated into the final model. Hospitals were also involved in decisions regarding patient exclusions.

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.**
Process of collaboration with hospitals and clinicians includes regular in person meetings, phone conferences, webinars. Hospitals receive quarterly reports of their submitted data results benchmarked to statewide averages and trended over time. The format and content of those reports were adjusted based on suggestions and needs of the hospital audiences.

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.
Feedback is obtained through the processes referred to in 4d1.2. for the advisory group. Hospitals and clinicians are also able to (and they do) contact us directly via email/phone/letters with feedback and comments.
Feedback to date on the risk adjustment model, variables, and performance has been positive from the clinical and hospital community. Hospitals appreciated the opportunity to comment on the model prior to finalization as well as contribute to the data variables for collection. Hospitals positively received the decision to exclude all transferred patients from their specific performance results. Some specialized hospitals (cancer centers) were concerned about whether the model adequately adjusts or considers their particular patient population and risks. Feedback on data collection suggests that abstraction remains a significant task/burden for hospitals.

4d2.2. **Summarize the feedback obtained from those being measured.**
The feedback has been positive with expected (from other public reporting initiatives in New York) concerns regarding whether this particular model, or any risk adjustment model, adequately adjusts for all the pertinent patient variables that can account for patient mix differences impacting mortality.

4d2.3. **Summarize the feedback obtained from other users**
Users identify the need to account for patient mix differences in the evaluation of outcomes such as mortality. While controversy and challenges exist regarding the use of formalized sepsis protocols, specific elements of sepsis bundles and some resuscitation interventions there has not been dispute that inpatient mortality remains an important outcome for both clinicians and patients/families. Having risk adjusted mortality rates allows users to identify those institutions for outreach who have demonstrated outstanding results. It is also the basis to better understand the impact of interventions, hospital characteristics associated with better or worse outcomes, whether there are subpopulations of patients with sepsis at higher or lower risk of mortality, or higher or lower likelihood of benefit/harm from interventions.

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.**
Several variables in the initial model were revised based on feedback from the advisory group. That included removal of ICU stay from the model (intervention variable potentially related to hospital resources and policies regarding ICU use) and the ´adding back´ in the model of cancer/lymphoma/leukemia variables, which did significantly impact some hospitals.

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)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0500: Severe Sepsis and Septic Shock: Management Bundle

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.**

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.)

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):** New York State Department of Health, Office of Quality and Patient Safety
Co.2 **Point of Contact:** Foster, Gesten, foster.gesten@health.ny.gov,518-486-6865-
Co.3 **Measure Developer if different from Measure Steward:** New York State Department of Health, Office of Quality and Patient Safety
Co.4 **Point of Contact:** Foster, Gesten, foster.gesten@health.ny.gov,518-486-6865-

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.

Foster Gesten, MD, and Marcus Friedrich, MD, New York State Department of Health: led sepsis initiative in NYS including development of regulations, review of sepsis protocols, development of data dictionary, contracting for statistical and clinical consultants and sepsis advisory group

Mitchell Levy, MD, Brown University: Chief clinical consultant for development of risk adjusted model, in collaboration with biostatisticians, New York State Department of Health, IPRO, and sepsis advisory group

Stanley Lemeshow, PhD,and Gary Phillips, Ohio State University: biostatisticians responsible for all analyses and model building associated with the risk adjusted mortality measure

Kathy Terry, PhD, IPRO: Project manager for sepsis data collection, data integrity, measure development and feedback reporting to hospitals

Edward Hannan,PhD, State University of New York at Albany, School of Public Health: consultant on risk adjustment methodology and modeling

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2017

Ad.3 **Month and Year of most recent revision:** 11, 2016

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

Ad.5 **When is the next scheduled review/update for this measure?** 09, 2017

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

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 #: 0500

Measure Title: Severe Sepsis and Septic Shock: Management Bundle

Measure Steward: Henry Ford Hospital

Brief Description of Measure: This measure focuses on adults 18 years and older with a diagnosis of severe sepsis or septic shock. Consistent with Surviving Sepsis Campaign guidelines, the measure contains several elements, including measurement of lactate, obtaining blood cultures, administering broad spectrum antibiotics, fluid resuscitation, vasopressor administration, reassessment of volume status and tissue perfusion, and repeat lactate measurement. As reflected in the data elements and their definitions, these elements should be performed in the early management of severe sepsis and septic shock.

Developer Rationale: A composite measure was developed given the dependencies the components have on one another. In addition, the components of the measure must be applied within specific time frames. The sequencing of the measure is such that the components could not stand alone unless certain preceding conditions had been met. In this way, treating the elements as a composite ensured assessment of a concerted strategy aimed at reducing mortality. The composite is more powerful that any individual application of the components in isolation from each other.

Numerator Statement: The number of patients in the denominator who received ALL of the following components (if applicable) for the early management of severe sepsis and septic shock: initial lactate levels, blood cultures, antibiotics, fluid resuscitation, repeat lactate level, vasopressors, and volume status and tissue perfusion reassessment.

- Within 3 hours of presentation of severe sepsis:
  - Measure initial lactate level
  - Draw blood cultures prior to antibiotics
  - Administer broad spectrum or other antibiotics
- Within 6 hours of presentation of severe sepsis:
  - Repeat lactate level (if initial lactate > 2 mmol/L)
- Within 3 hours of presentation of septic shock:
  - Administer 30 ml/kg crystalloid for hypotension or lactate ≥ 4 mmol/L
- Within 6 hours of presentation of septic shock:
  - Apply vasopressors (for hypotension that does not respond to initial fluid resuscitation) to maintain a mean arterial pressure (MAP) ≥ 65 mm Hg
  - Reasses volume status and tissue perfusion in the event of persistent hypotension (MAP <65 mm Hg) after initial fluid administration or initial lactate level ≥ 4 mmol/L
    - The clinician is no longer required to state the method of reassessment used (e.g. physical exam, bedside cardiovascular ultrasound, passive leg raising, CVP, ScV02 assessment). The clinician can attest that volume and perfusion reassessment has occurred, even without reference to the method used. This will meet the measure's volume and perfusion reassessment requirement. A provider may also opt to state their chosen method, but this is not required.

Denominator Statement: Inpatients age 18 and over with an ICD-10-CM Principal or Other Diagnosis Code of Sepsis, Severe Sepsis, or Septic Shock.

Denominator Exclusions: The following patients are excluded from the denominator:

- Severe sepsis is not present
- Patients Transferred in from another acute care facility
- Patients receiving IV antibiotics for more than 24 hours prior to presentation of severe sepsis.

- Patients with a Directive for Comfort Care or Palliative Care within 3 hours of presentation of severe sepsis
- Patients with an Administrative Contraindication to Care within 6 hours of presentation of severe sepsis
- Patients with an Administrative Contraindication to Care within 6 hours of presentation of septic shock
- Patients with a Directive for Comfort Care or Palliative Care within 6 hours of presentation of septic shock
- Patients with septic shock who are discharged within 6 hours of presentation
- Patients with severe sepsis who are discharged within 6 hours of presentation
- Patients with a Length of Stay >120 days
- Patients included in a Clinical Trial

Measure Type: Composite
Data Source: Imaging-Diagnostic, Laboratory, Other, Paper Records, Pharmacy
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Jun 07, 2012 **Most Recent Endorsement Date:** Nov 10, 2014

Composite Measure Construction: all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)
Component Measures (if endorsed or submitted for endorsement):

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**

Summary of prior review in 2012 and 2014:

- The developer provided the 2008 Surviving Sepsis Campaign guidelines with recommendations for initial resuscitation, measuring lactate, obtaining appropriate blood cultures, antibiotic therapy, fluid therapy, vasopressors and monitoring central venous pressure (CVP) and central venous oxygen saturation ($ScvO_2$).
- The developer provided a systematic review (SR) of the body of evidence supporting the processes of early management for patients with severe sepsis and septic shock. The SR included the grading of the evidence and the quality and consistency; no details on the quantity of the evidence was provided.
- In 2012, concerns were raised about the level of evidence supporting invasive monitoring of CVP and $ScvO_2$. The Infectious Disease Steering Committee acknowledged these concerns yet determined that the current evidence at the time was sufficient to warrant endorsement of the full bundle, and the measure was approved as specified. NQF received an appeal and the Consensus Standards Approval Committee (CSAC) upheld the measure's endorsement with the condition that NQF commit to an immediate re-evaluation of the measure upon release of new evidence from several ongoing studies including the Protocolized Care for Early Septic Shock (ProCESS)trial.
- In 2014, the Patient Safety Standing Committee conducted an *ad hoc* review based on a request from the American College of Emergency Physicians. The *ad hoc* review focused on the evidence supporting CVP and $ScvO_2$ and the new data from the ProCESS trial. See NQF-Endorsed Measures for Patient Safety (January 30,

2015) for complete summary.

- The ProCESS trial demonstrated no difference in mortality outcomes when using an invasive approach to monitoring CVP and $ScvO_2$ compared to usual care or protocolized care without invasive monitoring. The Committee noted that the new results from the ProCESS trial suggested that a mandate to measure CVP and $ScvO_2$ with an invasive line may not be necessary in all patients with severe sepsis and septic shock.
- Experts in support of maintaining these elements in the measure argued that additional trials (ARISE and PROMISE) were underway; however, these trials were smaller than the ProCESS trial and not performed in the U.S. In addition, these experts argued that the protocolized care and requirement for CVP and ScvO2 monitoring was particularly helpful in community hospitals, which were not included in the ProCESS trial.
- After extensive discussions and negotiations the measure developers, ProCESS trial investigators and speciality societies (including SCCM and ACEP) reached a compromise for an evidence-based replacement element – optional measurement of CVP and ScvO2, along with reassessment by other means (re-assess volume status and tissue perfusion after initial resuscitation and document findings).

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 evidence has been updated to support the changes to the measure since the last submission – see below.

This all-or-none composite measure assesses the proportion of facilities that measure initial lactate levels and repeat if elevated, obtain blood cultures, administer broad spectrum antibiotics, initiate fluid resuscitation, administer vasopressors, and reassess volume status and tissue perfusion for inpatients 18 years and older with a diagnosis of severe sepsis or septic shock.

- The developer included a diagram with the steps between the diagnosis of patients with severe sepsis or septic shock who receive all the elements of care in the sepsis bundle and the reduced risk of mortality.
- The developer provided a synthesis of the literature for the following updated components, which are based on the Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. ***[Note: Grading of recommendations for the components below are taken from the Sepsis and Septic Shock 2016 Guidelines]***
  - Measure lactate level; Repeat lactate if initial lactate is elevated *[weak recommendation, low quality of evidence[1]]*
  - Obtain cultures prior to antibiotics *[Best practice statement]*
  - Administer broad spectrum antibiotics *[strong recommendation, moderate quality of evidence]*
  - Administer 30 ml/kg crystalloid for hypotension or lactate ≥ 4 mmol/L *[strong recommendation, low quality of evidence]*
  - Apply vasopressors (for hypotension that does not respond to initial fluid resuscitation to maintain a mean arterial pressure ≥ 65 mmHg) *[strong recommendation, moderate quality of evidence]*
  - Reassess volume status and tissue perfusion *[Best practice statement]*
    - The developer provided a synthesis of the literature for some common practices used when reassessing volume status and tissue perfusion in the event of persistent hypotension (MAP <65 mm Hg) after initial fluid administration or initial lactate level ≥ 4 mmol/L. ***[Note: Clinician is no longer required to document the method used; attestation is sufficient]***
- The developer included the consistent reduction in mortality rates from 1992 – 2015 based on observational

[1] Per guideline authors, 'low' grade assigned to quality of evidence (5 RCTs) because 1) all studies were judged to be at high risk of bias due to lack of clarity of the intervention, therefore, we downgraded the quality of evidence by one level for risk of bias; 2) We downgraded the quality of evidence by one level for imprecision, the CI contained small benefit that was lower than the decision threshold; and 3) We assumed a mortality rate for patients with septic shock to be 40%.

studies of severe and septic shock in the United States, United Kingdom, and Australia and New Zealand. The developer also provided additional details about the type and number of studies conducted on sepsis mortality. The developer noted that, no studies (observational or otherwise), exist in the literature that fail to demonstrate a reduction in mortality when "sepsis bundles" employing all or nothing composite measures have been applied.

- The mortality rate for patients who received all applicable elements of care for the sepsis bundle (SEP-1) in CMS' Inpatient Quality Reporting (IQR) program was on average 8.5% lower compared to patients who did not receive all applicable elements of care.

Exception to evidence: N/A

Questions for the Committee:

- *The evidence provided by the developer is updated and consistent with the 2016 Guidelines for Management of Sepsis and Septic Shock. Does the Committee agree the evidence is sufficient and there is no need to re-vote on Evidence?*

Guidance from the Evidence Algorithm: Composite measure with systematic review (SR) and grading of the body of evidence (Box 3) → For each component, the Quantity/Quality/Consistency (QQC) of the body of evidence from the SR is available in the 2016 Sepsis and Septic Shock Guidelines or the Supplemental Digital Content 2 (Box 4) → For the composite measure, the SR concludes that there is moderate certainty that the net benefit is substantial OR moderate-high certainty the net benefit is moderate (Box 5b) → **Moderate**

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

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 provided the following performance rates for the composite measure as a whole and each component from CMS' Hospital Inpatient Quality Reporting (IQR) program from October 2015 to June 2016:

Overall composite measure rates:

	Oct-Dec 2015	Jan-Mar 2016	Apr-Jun 2016
# of hospitals	3,134	3,182	3,193
# of eligible cases	96,516	104,166	101,599
Overall performance rate	34.4	39.5	44.0
10th percentile	5.0	7.7	12.5
25th percentile	17.9	21.6	25.8
Median	31.0	36.1	41.7
75th percentile	45.8	51.3	57.1
90th percentile	60.0	66.7	71.4
Min, Max	0.0, 100.0	0.0, 100.0	0.0, 100.0
Average	32.6	37.1	41.9
Standard Deviation	21.1	21.9	22.9

Component rates categorized by 3 and 6 hour elements:

Population Description	Cases	Bundle Percentage
Initial Population Number of Sepsis Cases	**325,809**	**--**
Total Number of Excluded Sepsis Cases	*166,520*	--
Total Number of Eligible Sepsis Cases	**159,289**	**--**

Total Number of Passed Sepsis Cases	*64,051*	--
Total Number of Failed Sepsis Cases	*95,238*	--
Severe Sepsis 3 Hour Bundle Eligible Cases	**167,114**	--
Severe Sepsis 3 Hour Bundle Passes	*110,078*	*65.9%*
Severe Sepsis 3 Hour Bundle Failures	*54,618*	*32.7%*
Initial Lactate Level Failures	26,503	48.5%
Broad Spectrum or Other Antibiotic Administration Failures	20,951	38.4%
Blood Culture Collection Failures	18,772	34.4%
Severe Sepsis 6 Hour Bundle Eligible Cases	**90,385**	--
Severe Sepsis 6 Hour Bundle Passes	*53,475*	*59.2%*
Severe Sepsis 6 Hour Bundle Failures	*36,910*	*40.8%*
Repeat Lactate Level Failures	36,910	
Septic Shock 3 Hour Bundle Eligible Cases	**40,989**	--
Septic Shock 3 Hour Bundle Passes	*22,359*	*54.5%*
Septic Shock 3 Hour Bundle Failures	*18,630*	*45.5%*
Crystalloid Fluid Administration Failures	18,630	
Vasopressor Shock 6 Hour Bundle Eligible Cases	**8,177**	--
Vasopressor Shock 6 Hour Bundle Passes	*6,157*	*75.3%*
Vasopressor Shock 6 Hour Bundle Failures	*2,020*	*24.7%*
Vasopressor Administration Failures	2,020	
Focus Exam Shock 6 Hour Bundle Eligible Cases	**14,630**	--
Focus Exam Shock 6 Hour Bundle Passes	*3,801*	*26.0%*
Focus Exam Shock 6 Hour Bundle Failures	*9,935*	*67.9%*
Hemodynamic Choices Shock 6 Hour Bundle Eligible Cases	**10,829**	--
Hemodynamic Choices Shock 6 Hour Bundle Passes	*894*	*8.3%*
Hemodynamic Choices Shock 6 Hour Bundle Failures	*9,935*	*91.7%*

- The developer noted that the repeat volume and perfusion assessment data is broken down into focused exam and hemodynamic elements that are no longer required. No data is yet available on the new attestation strategy.
- The developer provided the chart below to describe the categorization of the 3 and 6 hour elements. The developer stated that for the purpose of this analysis the '30 ml/kg crystalloid fluid started' was grouped with the shock bundles, although technically only refractive hypotension qualifies for shock.

<table>
<tr><th rowspan="2">Required Action</th><th colspan="2">Severe Sepsis</th><th colspan="2">Septic Shock</th></tr>
<tr><th>3 hour Bundle</th><th>6 hour Bundle</th><th>3 hour Bundle</th><th>6 hour Bundle</th></tr>
<tr><td>Initial Lactate Collection</td><td>Yes</td><td colspan="3" rowspan="3">Must be completed within 3 hours of Severe Sepsis Presentation</td></tr>
<tr><td>Blood Culture Collection</td><td>Yes</td></tr>
<tr><td>Initial Antibiotic Started</td><td>Yes</td></tr>
<tr><td>Repeat Lactate Collection (if Initial Lactate > 2)</td><td colspan="2">Yes</td><td colspan="2">Must be completed within 6 hours of Severe Sepsis presentation</td></tr>
<tr><td>30mL/kg Crystalloid Fluids Started</td><td>N/A</td><td>N/A</td><td>Yes</td><td>Must be completed within 3 hrs of Septic Shock</td></tr>
<tr><td>Vasopressor Given (if ↓ BP persists)</td><td>N/A</td><td>N/A</td><td rowspan="2">Must be completed within 6 hrs of Septic Shock</td><td>Yes</td></tr>
<tr><td>Repeat Volume Status/ Tissue Perfusion Assessment</td><td>N/A</td><td>N/A</td><td>Yes</td></tr>
</table>

Disparities:

- The developer provided the following performance rates by ethnicity, gender and Medicare/non-Medicare:

	Oct-Dec 2015	Jan-Mar 2016	Apr-Jun 2016
Hispanic	34.53	39.8	44.23
Non-Hispanic	32.64	36.35	40.82
Females	33.82	39.22	43.28
Males	34.93	39.86	44.63
Medicare	34.9	40.07	44.57
Non-Medicare	33.32	38.47	42.76
Black	30.64	35.93	40.29
White	34.95	40.08	44.58
Other	34.95	40.01	43.82

- The performance rates provided for different age categories were similar (~34.0).

Questions for the Committee:

- *Does the composite measure demonstrate a quality problem in severe sepsis and septic shock care?*
- *Is a national performance measure still warranted?*
- *Are you aware of evidence that other disparities exist in this area of healthcare?*

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

1c. Composite - Quality Construct and Rationale
Maintenance measures – same emphasis on quality construct and rationale as for new measures.

1c. Composite Quality Construct and Rationale. The quality construct and rationale should be explicitly articulated and logical; a description of how the aggregation and weighting of the components is consistent with the quality construct and rationale also should be explicitly articulated and logical.

- This is an all-or-none composite measure. Patients with severe sepsis or septic shock must meet all of the eligible components in the composite.
  - The components include: lactate collection, delivery of broad spectrum antibiotics, obtaining blood cultures, delivering resuscitation fluids, applying vasopressors as needed, reassessing volume and perfusion status and repeating lactate values.
- The developer stated that this composite measure was developed due to the dependencies of the

components on one another and because the components must be applied within specific time frames.
- The components could not stand alone unless certain preceding conditions have been met. Per the developer, the composite ensures a strategy aimed at reducing mortality.
- The developer stated that the components are "aggregated both in time with 3 and 6 hour elements for severe sepsis and septic shock."
- The elements are equally weighted.

Questions for the Committee:

- *Are the quality construct and a rationale for the composite explicitly stated and logical?*
- *Is the method for aggregation and weighting of the components explicitly stated and logical?*

Preliminary rating for composite quality construct and rationale: ☒ High ☐ Moderate ☐ Low ☐ Insufficient

Committee pre-evaluation comments

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

Note: These comments were submitted prior to the updated measure information and preliminary analysis.

1. Importance to Measure and Report

1a. Evidence to Support Measure Focus

- This is a previously endorsed measure. It is a composite all or nothing process measure of elements of a "sepsis bundle." Abundant evidence is cited, most of it before the first review of this measure, but I believe it is overstated. For example, the Keystone project is listed as demonstrating improvement, but the authors conclude: "Participation in the Keystone Sepsis collaborative was unable to improve patient outcomes beyond concurrent trends. High bundle adherence hospitals had significantly greater improvements in outcomes, but further work is needed to understand these findings." Also, this pivotal study showing denominator inflation leads to overly optimistic conclusions about trends: https://www.ncbi.nlm.nih.gov/pubmed/27909698. The statement that observational studies provide the same level of evidence as RCTs is highly debatable given the Women's Health Study and other literature. There is a long section on 'possible options' that seems irrelevant to me, and is potentially self-serving. There is no doubt that correct early antibiotic treatment is important in reducing mortality, but since almost no patients receive anti-fungals empirically (especially in the ED), the data based on candidemia is misleading. I am not sure what Figure 1 shows - looks like a monotonic decrease in mortality over time without a discernible break in the slope that might be attributed to the interventions reviewed here. In short, there is a lot of evidence to suggest that these process measures are linked to outcomes, but it is overstated in the application.
- The bundle process is supported by the evidence; however, it is important to account for the ProCESS trial's results (which are cited but not fully addressed) as it's unclear if the whole bundle is what is at play or certain individual components.
- The evidence relates well and applies directly to the measure. Data on outcome related to SSC bundles shows decreased mortality in a variety of organizations. Some elements of the measure may require new definition as recommended in recent 2016 SSC, namely severe sepsis. Additionally, clarification may be needed for timing of process to manage infections (1 h vs 3 h) and better definition of what may be acceptable delay in blood culture collection (i.e. 45 minutes?).
- This is a composite process measure and has great complexity- it is not entirely clear what serves as a trigger to initiate the measure within the 3 and 6 hour windows
- There is evidence supporting many individual elements of the measure, and the measure in general, however there are NO studies that show the impact (or risk) of 100% compliance. This raises concern about confounding my indication and unmeasured confounders. This also raises concern given that the average compliance in the studies is quite low. The question is what happens to outcomes of the 50%+ of people in whom their physician has chosen not to give them this treatment. We know that with marginal increased compliance outcomes are better, but what about the top 25% (who were least likely to get the bundles) - is it safe for them? The issue of the evidence extends to some of the individual recommendations such as fluid bolus in ALL patients (including those with HF) - the studies cited are all retrospective reviews, with no randomized data to support that it is safe in all HR failure patients - instead this only confirms that in the HR failure patients in whom their MDs were comfortable with them getting the fluid, they did ok.

1b. Performance Gap

- There are good data on persisting gaps and disparities based on the IQR program analysis to-date.
- Yes.

- Performance data were provided from studies and CMS (may need clarification to confirm that is the source of the data). Data was provided for age, gender, ethnicity and race. Clarification that patients from other races had similar rate of receiving all applicable elements compared to whites should state "except for blacks". Specific data as it relates to blacks vs other races was not provided. References for the rational for the measure should be updated to reflect the most recent guidelines and best practice recommendations.
- Several of the measures have solid support in the literature: measurement of lactate, blood cultures and administration of broad spectrum antibiotics.
- The measures of perfusion and resuscitation are problematic and less well accepted. The measure suggests that CVP measurement (including ScvO2) be routine - the evidence for this is not as strong.
- Yes - there is marked variability in care. So much so, that it seems that a national measure (IF measurable!) could be useful. There is also some evidence of disparities, although (while statistically significant) these are very small. It would also be critical to know if this is controlled for by institution. I suspect that certain institutions have higher proportions of patients receiving all applicable elements of clinical care, as well as higher/lower rates of certain demographic populations.

1c. Composite Performance Measure –Quality Construct

- Yes
- Yes
- Overall area of quality is stated clearly (need to be updated to delete "severe" sepsis). As this is an all-or-none measure, statement of "as needed" for applying vasopressors may need clarification. Should a 1 hour time point be considered for some of the elements (lactate, blood culture and antibiotics).
- This is stated in general terms. The logic for including these variables, which occur in a limited time span and are sequential, is clear. However, I am less impressed by review of the effect of each component and that greater effect of lumping them together. It's hard to tease out whether these should be weighted and what the attributable effect is of each variable in the composite measure.

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): Imaging-Diagnostic, Laboratory, Other, Paper Records, Pharmacy

Specifications:

- The is a facility-level, all-or-none composite measure.
- Per the developer, the measure has undergone at least three rounds of updates since the last endorsement. See details of all changes/refinement to specifications here.
- The denominator includes inpatients age 18 and over with an ICD-10-CM Principal or Other Diagnosis Code of Sepsis, Severe Sepsis, or Septic Shock.
- The numerator includes the number of patients in the denominator who received **ALL** of the following components (if applicable) for the early management of severe sepsis and septic shock:
    - For **severe sepsis**: measure initial lactate level, draw blood cultures prior to antibiotics, administer broad spectrum or other antibiotics within **3 hours** of presentation. Repeat lactate level (if initial lactate > 2 mmol/L) within **6 hours** of presentation.
      For **septic shock**: administer 30 ml/kg crystalloid for hypotension or lactate ≥ 4 mmol/L within **3 hours** of presentation. Apply vasopressors (for hypotension that does not respond to initial fluid resuscitation) to maintain a mean arterial pressure (MAP) ≥ 65 mm Hg and reassess volume status and tissue perfusion in the event of persistent hypotension (MAP <65 mm Hg) after initial fluid administration or initial lactate level ≥ 4 mmol/L within **6 hours** of presentation.

- *Note: The measure no longer requires a provider to state the method of reassessment used (e.g. physical exam, bedside cardiovascular ultrasound, passive leg raising, CVP, ScV02 assessment). The clinician can attest that volume and perfusion reassessment has occurred, even without reference to the method used. This will meet the measure's volume and perfusion reassessment requirement. A provider may also opt to state their chosen method, but this is not required.*

- The denominator exclusions include:
  - Severe sepsis is not present
  - Patients Transferred in from another acute care facility
  - Patients receiving IV antibiotics for more than 24 hours prior to presentation of severe sepsis
  - Patients with a Directive for Comfort Care or Palliative Care within 3 hours of presentation of severe sepsis
  - Patients with an Administrative Contraindication to Care within 6 hours of presentation of severe sepsis
  - Patients with an Administrative Contraindication to Care within 6 hours of presentation of septic shock
  - Patients with a Directive for Comfort Care or Palliative Care within 6 hours of presentation of septic shock
  - Patients with septic shock who are discharged within 6 hours of presentation
  - Patients with severe sepsis who are discharged within 6 hours of presentation
  - Patients with a Length of Stay >120 days
  - Patients included in a Clinical Trial
- ICD-10-CM codes for the denominator are provided.
- The measure is not risk-adjusted or stratified.
- The calculation algorithm with 141 individual steps is provided; however, there is no information in the algorithm describing compliance with the components for the 3 and 6 hour time frames.
- Instructions for monthly and quarterly sampling are included.

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 is 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 2012, the developer conducted a signal-to-noise analysis of the individual bundle elements and overall bundle reliability and composite measure reliability by site for 165 hospitals and 15,022 patients from January 2005 – March 2008. *[Reliability must be demonstrated for the composite measure score; reliability of the individual component measures is not sufficient – effective 2013]*

Describe any updates to testing: see updated composite measure score level testing below

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 a random sample of SEP-1 chart-abstracted data submitted to CMS as part of the Hospital Inpatient Quality Reporting (IQR) program from October 2015 to June 2016. The sample included 302,281 cases in the denominator (after exclusions) and 119,048 cases in the numerator from 3,134 to 3,193 hospitals (depending on the quarter).
- The developer used a beta-binomial model to assess the signal-to-noise ratio. A reliability of 0.0 implies that all

the variability in a measure is attributable to measurement error. A reliability of 1.00 implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one provider from another. This is an appropriate test for measure score reliability. A reliability of 0.70 is generally considered a minimum threshold for reliability.
    - The median reliability score was calculated including all facilities and facilities with a minimum of 10 eligible cases (more than 86% of reporting facilities)

Results of reliability testing:

- The developer provided the overall reliability score for the composite measure for each quarter:
    - October 2015 – December 2015: 0.92 (CI 0.41 – 1.00)
    - January 2016 – March 2016: 0.93 (CI 0.47 – 1.00)
    - April 2016 – June 2016: 0.93 (CI 0.42 – 1.00)
- The developer stated that the facilities with a minimum of 10 cases per quarter had similar reliability scores with narrowing confidence intervals.

Questions for the Committee:

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

Guidance from the Reliability Algorithm: Precise specifications (Box 1) → Empirical reliability testing conducted using statistical tests with the measure as specified (Box 2) → Reliability testing conducted with composite measure score (Box 4) → Appropriate method used for assessing the proportion of variability due to real differences among measure entities (Box 5) → Moderate certainty or confidence that the performance measure scores are reliable based on the reliability statistic and scope of testing (number of measured entities and representativeness) (Box 6b) → **Moderate**

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:

- *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:

- In 2012, the developer assessed the validity of the measure score by testing the hypothesis that those with higher scores on the composite performance measure should have a lower score on a risk-adjusted mortality measure. The developer reported that hospital mortality was reduced by 10% for patients that were compliant with the composite measure.

Describe any updates to validity testing: see measure score validity testing 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 methods:

- The dataset included 159,289 patients from 3,134 to 3,193 hospitals (depending on the quarter) who were eligible for the SEP-1 measure from October 2015 to June 2016.
- The developer performed a Chi-Square Test of Association and Equal Proportions between two categorical variables: Measure Outcome (Failed or Passed) and Mortality Result (Died or Survived). The developer also included an analysis of pass rates and mortality rates by percentiles and a two-proportions z-test. The z-test determines if there is a statistically significant difference in mortality rates between percentiles. *These methods are appropriate for empirically assessing the validity of the composite measure score.*
- The developer provided patient-level data element validity testing – *this method is not appropriate for composite measures.*
- NQF composite performance measure evaluation guidance (2013) states that validity testing is directed toward the inferences that can be made about accountable entities on the basis of their performance measure scores. For the purposes of endorsing composite performance measures, validity testing of the constructed composite performance measure score is more important that validity testing of the component measures. Even if the individual component measures are valid, the aggregation and weighting rules for constructing the composite could result in a score that is not a true reflection of quality (p. 12-13).

Validity testing results:

- The developer provided the sepsis mortality analysis and the Chi-Square Test of Association and Equal Proportions below:

Sepsis Mortality Analysis

Population Description	Cases	Total Percentage	Total Deaths	Total Deaths Percentage
Initial Population Number of Sepsis Cases	**325,809**	**--**	**81,587**	**25.0%**
Total Number of Excluded Sepsis Cases	*166,520*	*51.1%*	*38,624*	*23.2%*
Total Number of Eligible Sepsis Cases	**159,289**	**48.9%**	**42,963**	**27.0%**
Total Number of Passed Sepsis Cases	*64,051*	*40.2%*	*14,039*	*21.9%*
Total Number of Failed Sepsis Cases	*95,238*	*59.8%*	*28,924*	*30.4%*

Chi-Square Test of Associate and Equal Proportions

Population	Chi-Square	P-Value	Risk Ratio	Lower 95% Confidence Limit	Upper 95% Confidence Limit
Total Severe Sepsis Eligible Cases	1388.8163	<.0001	1.3856	1.3616	1.4101

- The results of the sepsis mortality analysis demonstrate that **30.4%** of the total number of 'Failed Sepsis Cases' died (at discharge and up to 30 days after discharge) compared to **21.9%** of the total number of 'Passed Sepsis Cases'.
- A Risk Ratio higher than 1.0 with a significant p-value, would indicate that there is a higher risk of dying when a case fails the measure compared to when a case passes the measure.
    - The Risk Ratio can be used as an actual ratio and it can be said that with 95% confidence, cases that fail the measure have 1.36 to 1.41 times the risk of dying compared to cases that pass the measure.
    - The Risk Ratio can also be used as a percentage and be said that with 95% confidence, cases that fail the

measure have a 36% to 41% increase in risk of dying compared to cases that pass the measure.

- The sepsis rate comparisons analyses provided by the developer demonstrate a negative association between pass rates and mortality rates from October 2015 to June 2016. See additional analyses for each quarter in spreadsheet uploaded to ShP.
- The two-proportion z-test demonstrates that four of the percentile comparisons have a statistically significant difference between mortality rates at a significance level of 0.05. Three additional percentile comparisons are fairly close to a statistically significant difference between mortality rates at a significance level of 0.10.
- The developer also provided the mortality rate for patients who received all applicable elements of care for the composite measure (passed sepsis cases) and those who did not (failed sepsis cases) for each quarter – the mortality rate was on average 8.5% lower compared to patients who did not receive all applicable elements of care.

	Severe Sepsis and Septic Shock Mortality Rate		
Description	**2015 Q4**	**2016 Q1**	**2016 Q2**
Did not Meet Guidelines for SEP-1	29.6%	31.8%	29.7%
Met Guidelines for SEP-1	21.3%	23.0%	21.4%
Absolute Reduction Rate	8.3%	8.8%	8.3%
Relative Reduction Rate	28.04%	27.7%	27.9%
	Potential Preventable Deaths		
	2,783	2,864	2,411

- Per the developer, the mortality rates demonstrate the potential impact of the measure in the number of preventable deaths as a result of the measure. The relative mortality reduction observed with the introduction of the measure is similar to that seen with quality initiatives over the last 15 years.

Questions for the Committee:

- *The prior testing demonstrated good measure score validity based on the previous specifications. The specifications have changed and the developer has provided updated validity testing of the composite measure score since the measure has been implemented. Does the Committee think there is a need to re-discuss and re-vote on validity?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality related to patients with severe sepsis and septic shock can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer provided the following analysis of exclusions from 3,134 - 3,193 facilities (depending on the quarter) from October 2015 to June 2016:

Algorithm	Data Element Name	Cases	Percent
	Total Number of Excluded SEP-1 Cases	339,678	100.00%
Step-04	Severe Sepsis Present is not present	245,737	72.34%
Step-03	Transfer from Another Hospital	62,502	18.40%
Step-22	Broad Spectrum Antibiotic Time > 24 hours	13,112	3.86%
Step-07	Directive for Comfort Care, Severe Sepsis	9,920	2.92%

Step-02	Administrative Contraindication to care, Severe Sepsis and Septic Shock	6,007	1.77%
Step-42	Directive for Comfort Care, Septic Shock	1,256	0.37%
Step-45	Shock Expired Time < 6 hours	620	0.18%
Step-11	Sepsis Expired Time < 3 hours	524	0.15%

- The developer states that "the number exclusions are not significant enough to unfairly distort measure performance results and potentially negatively affect the reliability fo the measure because the vast majority of the exclusions are cases where severe sepsis is not present and should not be analyzed."
- The developer also states that there is a strong clinical rationale and previous precedent across other IQR measures to exclude cases with the above exclusions.

- Exclusions analysis did not include the following exclusions and their potential impact on the measure results:
  - Patients with a Length of Stay >120 days
  - Patients included in a Clinical Trial (not yet implemented)

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

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)*:

- The developer included the facility-level overall performance rates provided for performance gap and noted that the measure was able to detect facilities with better- and worse-than-average performance.
- The developer also stated that reporting a mean provides outlying facilities an opportunity to identify underperformance related to implementing applicable elements of clinical care in patients with severe sepsis and septic shock.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- The developer states that missing data is not a concern for this measure because the algorithm rejects cases and does not allow submission in instances where there is missing data for a data element.

Guidance from the Validity Algorithm: Specifications consistent with the evidence provided in support of the measure (Box 1) → Most of the potential threats to validity that are relevant to the measure empirically assessed (Box 2) → Empirical validity testing conducted using the measure as specified (Box 3) → Validity testing conducted with computed performance measure score (Box 6) → The method described was appropriate for assessing the relationship between the performance on this measure and mortality (Box 7) → Moderate certainty or confidence that the performance measure scores (from the measure as specified) are a valid indicator of quality based on the results (significance and strength) and scope of testing (number of representativeness) and analysis of potential threats (Box 8b) → **Moderate**

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

2d. Composite measure: empirical analysis supports construction

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

- The developer provided the following analysis demonstrating the contribution of each component to the composite score (e.g., frequency of failure on each component):

Population Description	Cases	Bundle Percentage
Initial Population Number of Sepsis Cases	**325,809**	**--**
Total Number of Excluded Sepsis Cases	*166,520*	--
Total Number of Eligible Sepsis Cases	**159,289**	**--**
Total Number of Passed Sepsis Cases	*64,051*	--
Total Number of Failed Sepsis Cases	*95,238*	--
Severe Sepsis 3 Hour Bundle Eligible Cases	**167,114**	**--**
Severe Sepsis 3 Hour Bundle Passes	*110,078*	*65.9%*
Severe Sepsis 3 Hour Bundle Failures	*54,618*	*32.7%*
Initial Lactate Level Failures	26,503	48.5%
Broad Spectrum or Other Antibiotic Administration Failures	20,951	38.4%
Blood Culture Collection Failures	18,772	34.4%
Severe Sepsis 6 Hour Bundle Eligible Cases	**90,385**	**--**
Severe Sepsis 6 Hour Bundle Passes	*53,475*	*59.2%*
Severe Sepsis 6 Hour Bundle Failures	*36,910*	*40.8%*
Repeat Lactate Level Failures	36,910	
Septic Shock 3 Hour Bundle Eligible Cases	**40,989**	**--**
Septic Shock 3 Hour Bundle Passes	*22,359*	*54.5%*
Septic Shock 3 Hour Bundle Failures	*18,630*	*45.5%*
Crystalloid Fluid Administration Failures	18,630	
Vasopressor Shock 6 Hour Bundle Eligible Cases	**8,177**	**--**
Vasopressor Shock 6 Hour Bundle Passes	*6,157*	*75.3%*
Vasopressor Shock 6 Hour Bundle Failures	*2,020*	*24.7%*
Vasopressor Administration Failures	2,020	
Focus Exam Shock 6 Hour Bundle Eligible Cases	**14,630**	**--**
Focus Exam Shock 6 Hour Bundle Passes	*3,801*	*26.0%*
Focus Exam Shock 6 Hour Bundle Failures	*9,935*	*67.9%*
Hemodynamic Choices Shock 6 Hour Bundle Eligible Cases	**10,829**	**--**
Hemodynamic Choices Shock 6 Hour Bundle Passes	*894*	*8.3%*
Hemodynamic Choices Shock 6 Hour Bundle Failures	*9,935*	*91.7%*

- The developer noted that the repeat volume and perfusion assessment data is broken down into focused exam and hemodynamic elements that are no longer required. No data is yet available on the new attestation strategy.
- The developer provided the chart below to describe the categorization of the 3 and 6 hour elements. The developer states that for the purpose of this analysis the '30 ml/kg crystalloid fluid started' was grouped with the shock bundles, although technically only refractive hypotension qualifies for shock.

<table>
<tr><th rowspan="2">Required Action</th><th colspan="2">Severe Sepsis</th><th colspan="2">Septic Shock</th></tr>
<tr><th>3 hour Bundle</th><th>6 hour Bundle</th><th>3 hour Bundle</th><th>6 hour Bundle</th></tr>
<tr><td>Initial Lactate Collection</td><td>Yes</td><td colspan="3" rowspan="3">Must be completed within 3 hours of Severe Sepsis Presentation</td></tr>
<tr><td>Blood Culture Collection</td><td>Yes</td></tr>
<tr><td>Initial Antibiotic Started</td><td>Yes</td></tr>
<tr><td>Repeat Lactate Collection (if Initial Lactate > 2)</td><td colspan="2">Yes</td><td colspan="2">Must be completed within 6 hours of Severe Sepsis presentation</td></tr>
<tr><td>30mL/kg Crystalloid Fluids Started</td><td>N/A</td><td>N/A</td><td>Yes</td><td>Must be completed within 3 hrs of Septic Shock</td></tr>
<tr><td>Vasopressor Given (if ↓ BP persists)</td><td>N/A</td><td>N/A</td><td rowspan="2">Must be completed within 6 hrs of Septic Shock</td><td>Yes</td></tr>
<tr><td>Repeat Volume Status/ Tissue Perfusion Assessment</td><td>N/A</td><td>N/A</td><td>Yes</td></tr>
</table>

Questions for the Committee:

- *Do the components add value to the composite measure?*
- *Does the analysis support the aggregation and weighting rules of the composite measure?*
- *Are the objectives of parsimony and simplicity achieved while supporting the quality construct?*

Preliminary rating for composite construction: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

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

Note: These comments were submitted prior to the updated measure information and preliminary analysis.

2a1. & 2b1. Specifications

- This seems pretty clear to me based on my limited experience in specifications for NQF measures. However, I do have some concerns about adjudication of self reported "attestations" by clinicians (e.g., for antibiotic choice, fluid resuscitation, etc.) This will need to be discussed in the committee meeting. The efforts to reduce documentation and chart abstraction burden may be leading to more subjectivity, even though these are laudable goals
- None
- Main concern is the long list of steps in the logic or calculation algorithm. How is step 20 different from step 22; 26 vs 28, 35 vs 37, 46 vs 48, 60 vs 6265 vs 67, 72 vs 74, 39 vs 140, ect. Was acceptable blood culture delay defined?
- The data elements that trigger initiation of the measure (a rigorous definition of sepsis/septic shock) need clear definition within the measure.
- Interestingly, the overall reliability of the measure is stated as excellent (.92-.93), however the this is surprising given the marked issue with validity of some of the foundational elements (like time zero) that other elements of the composite measure are predicated on. Although each data element is clearly defined, there are multiple ways to determine that something meets criteria, and given the requirement to time stamp this likely is what results in poorly reproducible data.

2a2. Reliability Testing

- I did not see extensive discussion of inter- or intra observer reliability, particularly for aspects of the measure that may be subjective
- Yes
- Yes
- Reliability of the overall measure was tested, however given the complexity of the measure, reliability testing of each element should have been tested (as it was for validity testing).

2b2. Validity Testing

- I am leaning towards saying we do have enough data because the measure has been used widely and extensively, and some revisions have been made on the basis of the data. However, as noted previously, the linkage between process and outcome measures may be weaker than claimed. In particular, given the authors' statement that mortality rises precipitously beyond one hour for antibiotic treatment, it is not clear why a 3 hours threshold was chosen. And in reality, lactate is not generally available when antibiotics are chosen and given, so statements about its utility in that regard are questionable (if I understood the authors' statements correctly). It also is critical to note that no balancing measure is required or even mentioned, and we know that overdiagnosis of sepsis and over-administration of antibiotics has been a problem with past measures. The authors talk a lot about de-escalation but do not provide this as a balancing measure either. Similarly, there is little about overhydration of elderly patients with fragile cardiovascular systems. It seems problematic to me that the same measures are being applied to EDs, regular wards, and ICUs as these contexts are very different. It would have been desirable to stratify the data and discuss this issue.
- I think the ProCESS trial has to be incorporated more fully
- Validity was tested with an adequate number of entities and patients to generalized. As the majority of elements had agreement of <90% and would require further validation to confirm that they are useful indicator of quality (especially those less than 50%)
- See above - concern that the results of the validity testing show that 72.7% of elements are not valid (lower than 90% agreement).

2b. 3-7 Threats to Validity

- Exclusions seem reasonable in general, although I think patients with VAD are so different that they probably should be excluded. I am not sure where to state this in the form, but there is little mention of other causes of sepsis syndrome, such as influenza, and how antibiotic treatment would be evaluated in such patients. I don't see viruses or non-infectious sepsis mentioned in the coding section. The issue of "severe sepsis" is not carefully discussed, although I am sure criteria for this exist. In general, we know that there has been overcoding of sepsis and denominator inflation, so this threat needs to be discussed. Fortunately, risk adjustment is a relatively minor issue here since adherence to the composite measure should happen for all patients. However, the effect modification of the bundle on the desired outcome would be affected by case mix, and this is worth discussing. Case mix might also affect the nature of sepsis and conceivably could distort clinical decisions as to whether or not these interventions were indicated.
- No
- How is enough frequency define to warrant inclusion of an exclusion criteria (i.e. is sepsis expired time <3 hours necessary if only 0.15% of cases?)
- Exclusions are a concern related to abstractor burden. Of the 339678 cases included, 245737 were excluded because Severe Sepsis was not present. To make this determination requires a great degree of abstraction often, and is a high abstractor burden for a very low yield (aka 72.3% of cases that require detailed review are excluded).

  Of note, the median rate ranging from 31-41.7% is in keeping with compliance on studies looking at these bundles. Again, this raises concern that the safety of applying this care to 100% of patients has not been tested. There are likely populations that should be excluded that have not been identified/tested.

2d. Composite performance measures

- Please see previous comments. Composite makes sense, but the attributable effect and weight of each component are not clear
- Again, this is difficult because there is some debate over which elements of the bundle are most necessary.
- Yes - for sepsis the measure would have to be a composite measure. There is no weighting of the one element as more important than another (all or nothing).

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.

- Data elements are abstracted from a record by someone other than the person obtaining the original information (e.g. chart abstraction); some data elements are in defined fields in electronic sources.
- The developer noted that the measure is complex and requires data abstractors to "comb through documentation and interpret" difficult to find clinician documentation; however, the most recent updates to the measure should lessen documentation and abstraction burden.
- The developer also stated that preliminary efforts to convert this measure to an electronic measure within the HQMF/QDM framework was not feasible; there are no plans to develop an eMeasure at this time.
- The developer also states that the measure has gone through three updates where changes were made to ease abstractor burden and address issues related to data availability, missing data, and frequency of data collection.
- There are no fees or licenses required to use this 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

Note: These comments were submitted prior to the updated measure information and preliminary analysis.

3. Feasability

- Most of the required data probably are not in the EMR, and efforts to automate should continue. However, there is a lot of clinical judgment involved, documentation and attestations are burdensome, and chart abstraction is cumbersome even with the revisions. NLP should be explored. Has anyone asked the care team if they find the documentation burdensome?
- It's feasible
- This is highly dependent on hospital and laboratory information systems and data may not always be easily retrieved.
- Feasibility is a significant concern with this measure as specified. The measure requires extensive documentation of parameters that do not currently exist in discrete fields in the EHR such as bedside cardiovascular ultrasound, passive leg raise and focused exam parameters including skin finding and capillary refill. Any measures that are present in only free text documentation will require time consuming abstraction by clinical reviewers.The determination of the result of response to fluid administration will also be very difficult to determine and abstract. The complexity of the measure is such that it may only be successfully measured with template notes that are used when the septic patient is identified.
- This is a complex composite measure which is made up of elements which, although routinely recorded, are recorded variably in the EMR. The difficulties regarding data collection likely results in significant time and cost associated with collection this measure. Additionally, certain elements are questionably appropriate for abstractors (such as the chronicity of underlying conditions which makes their presence an appropriate exclusion).

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**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- Currently used in CMS' Hospital Inpatient Quality Reporting (IQR) Program for acute care hospitals nation-wide. Across the three quarters of available data, between 3,134 and 3,193 providers submitted data, which represents more than 95% of eligible providers nationwide.
- The measure is not currently publicly reported, but will be added to the Hospital Compare website at a date to be determined. Due to the complexity of the measure specifications, CMS desires to review and analyze the data prior to making it publicly available. There were also several updates to the specifications based on stakeholder feedback, and CMS wanted to assure stability of the specifications before public reporting.

Improvement results:

- The developer reports median performance across three quarters, demonstrating improvement from 31% in Q4 2015, to 36.1% in Q1 2016, and 41.7% in Q2 2016.

Unexpected findings (positive or negative) during implementation:

- None reported.

Potential harms:

- None reported.

Vetting of the measure:

- Per the developer, certain updates have been made do ease abstractor burden and mitigate issues related to availability of data (vital signs for vital signs review), missing data, and frequency of data collection (Clinician attestation to suffice requirements for five data elements). All of these relevant updates have positively impacted the time and cost of data collection for providers submitting data.
- No specific information included on the performance results, data, and assistance with interpreting the measure results provided to facilities but typically CMS provides facilities participating in IQR quarterly reports.

Feedback:

- On October 6, 2016, CMS provided the following update on qualitynet.org:
  The Centers for Medicare & Medicaid Services (CMS) updated the Severe Sepsis and Septic Shock: Management Bundle (SEP-1) measure specifications several times in response to newly published evidence. As a result, CMS will not score the SEP-1 measure validation for Hospital Inpatient Quality Reporting (IQR) Fiscal Year (FY) 2018. CMS is also postponing the public reporting of the SEP-1 measure on *Hospital Compare* until it is confident that it has valid data that reflects hospitals' performance.

  Ninety-nine percent of participating hospitals submitted SEP-1 measure data for Fourth Quarter 2015 and First Quarter 2016 within the submission timeframe. CMS is confident the specifications manual published in July 2016 for discharges beginning January 1, 2017, contains the stabilized SEP-1 measure specifications, and CMS will continue to analyze the data it receives to determine when the data will be used for public reporting.

  Validation plans for sepsis measure data
  Starting with Fourth Quarter 2015 data, CMS requested medical records for the SEP-1 measure from the hospitals selected for Hospital IQR Program validation. The scores for sepsis measure validation will not be used to calculate the confidence interval for the IQR validation; therefore, the validation of measure data from Fourth Quarter 2015, First Quarter 2016, and Second Quarter 2016 will not impact the Annual Payment Update (APU) of hospitals for FY 2018. Hospitals are still required to submit SEP-1 data. CMS will provide guidance on when the measure will be used for validation in future communications.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*
- *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

Note: These comments were submitted prior to the updated measure information and preliminary analysis.

4. Usability and Use

- The measure has been used extensively in the real world, but unless I missed it, the authors have not explored how it is RECEIVED in the real world by clinicians and what it adds to daily work. As noted above, there is a regrettable lack of attention to unintended consequences.
- Public data on compliance with the bundle and mortality rates for individual hospitals could help drive performance as well as create incentive programs for 3rd party payers and also lead to centers of excellence -- but again this is premised on the power of the entire bundle.
- Not being publicly reported currently.
- How will the measure be reported? Given the number of components within the measure, will compliance be reported as 100% adherence to all components? What about lower levels of adherence- how will that be reported? Are all elements of the measure of equal therapeutic value?
- The measure is not yet being publicly reported. CMS is collecting it, there is discussion of hospital compare, but no plans at the moment given concerns about reliability and validity.

Criterion 5: Related and Competing Measures

Related or competing measures

- 3215: Adult Sepsis Mortality Outcome Measure

Harmonization

- Measure #3215 is newly submitted to the Infectious Disease project. The related and competing discussion for these measures will take place during the in person meeting in March.

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 candidate 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:

Pre-meeting public and member comments

- The American Medical Association (AMA) appreciates the opportunity to comment prior to the NQF Infectious Disease Standing Committee's review. We strongly support the modifications that the developer made to the measure, #500, Severe Sepsis and Septic Shock: Management Bundle. Specifically, the additional requirement requiring documentation by a clinician to confirm severe sepsis addresses our concern that the previous denominator was too broadly defined and could have unintended negative consequences by including patients inappropriately. We also appreciate the additional enhancements made to the measure to capture individual patient circumstances, including capturing appropriate delays in obtaining blood cultures prior to antibiotic administration and the additional data elements to better identify those patients

for whom crystalloid fluid administration is suitable. We would request one clarification in the measure specifications regarding one of the exclusions. Specifically, the denominator exclusions in S.8 include the following element, Patient included in clinical trial; yet, S.3.2, where changes to the measure are outlined, states that this exclusion has been removed. It would be helpful to have the developer clarify this discrepancy prior to releasing the measure for public and member comment. We look forward to monitoring the discussion of the measure and the opportunity to comment during the review and endorsement process.

- The Federation of American Hospitals ("FAH") appreciates the opportunity to comment prior to the NQF Infectious Disease Standing Committee's review. FAH believes that effective and timely treatment of severe sepsis and septic shock in patients is vital and we support the intent of the measure. FAH appreciates the modification to the measure to allow for administration of targeted antibiotics when the causative organism and susceptibility are known. FAH encourages the developer to continue to monitor the measure to ensure that this additional flexibility addresses the ongoing efforts of effective antibiotic stewardship and minimizing the potential for overuse. The FAH also asks for clarification of one of the measure specifications. Specifically, the denominator exclusions in S.8 include the following element, "Patient included in clinical trial"; yet, S.3.2, where changes to the measure are outlined, states that this exclusion has been removed. The FAH strongly encourages the developer to clarify this discrepancy prior to releasing the measure for public and member comment.

The FAH remains extremely concerned that the sepsis measure continues to be overly complex and burdensome to collect, and, therefore, hampers the ability of hospitals to appropriately evaluate their performance on this measure since many of the data elements required for this measure can be captured only through chart abstraction. This complexity is evidenced by the validity testing results where forty data elements achieved less than 90% percent agreement between the data abstracted by hospitals and data abstracted by independent medical abstractors. FAH believes that the validity results demonstrate the inherent problems with implementing a measure with more than 140 data elements. In addition, even though hospitals collected the data for the measure since the last quarter of 2015 and the Centers for Medicare & Medicaid Services (CMS) now has a full year of data, CMS has not yet publicly reported results due to concerns with data quality. The FAH asks that the committee consider these findings during the measure evaluation. While the FAH understands there are limitations to what can be collected in electronic health records currently, we strongly urge the developer to revise the measure to enable electronic data capture. Developing an eMeasure in this area would further ensure that the relevant information is available for use at the point of care and facilitate communication with providers at the next level of care. Continuing to maintain and endorse a measure that requires manual abstraction with this many elements should not be viewed as a long-term solution given the current environment of promoting electronic data capture.

EVIDENCE FROM PREVIOUS SUBMISSION

NATIONAL QUALITY FORUM

Measure missing data in Composite 2.0 from Composite 1.0

NQF #: 0500 **NQF Project:** Infectious Disease Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Extent to which the specific measure focus is important to making significant gains in health care quality (safety, timeliness, effectiveness, efficiency, equity, patient-centeredness) and improving health outcomes for a specific high impact aspect of healthcare where there is variation in or overall poor performance.

Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.
(composite measure evaluation criteria)

1c.1 **Structure-Process-Outcome Relationship** *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The foci of this composite are the processes of early management for patients with severe sepsis and septic shock. All bundle elements are associated with improved outcomes for severe sepsis and septic shock patients including mortality and length of stay and have been consistently observed with implementation of early best practice intervention strategies.

1c.2-3 **Type of Evidence** *(Check all that apply)*:
Clinical Practice Guideline

Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Exclusions Justified A) Patients with advanced directives for comfort care are excluded.
B) Clinical conditions that preclude total measure completion should be excluded (e.g. mortality within the first 6 hours of presentation as defined above in 2a1.1).
C) Patients for whom a central line is clinically contraindicated (e.g. coagulopathy that cannot be corrected, inadequate internal jugular or subclavian central venous access due to repeated cannulations).
D) Patients for whom a central line was attempted but could not be successfully inserted.
E) Patient or surrogate decision maker declined or is unwilling to consent to such therapies or central line placement.

Please note that the exclusions are highly intuitive and reasonable. Thus, imagining a world for testing purposes, where patients who did not consent for lines received them or who wished to be made comfort measures were treated aggressively is wholly unlikely. Such a study most likely could not be conducted due to appropriate IRB constraints.

1c.5 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)***:**
The measure focus is on adults 18 years and older with a diagnosis of severe sepsis and septic shock. Consistent with Surviving Sepsis Campaign guidelines, it recommends measurement of lactate, obtaining blood cultures, administering broad spectrum antibiotics, fluid resuscitation, repeat lactate measurement for lactate clearance, and measuring central venous pressure (CVP) and central venous oxygen saturation (ScvO2). The evidence cited for all components of this measure is directly related to decreases in organ failure, overall reductions in hospital mortality, length of stay, and costs of care. For more information, please see attachment entitled NQF 0500 Tables and Forest Plots under the section "Scientific Acceptability".

1c.6 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles*)**:** The SSC guidelines support for this measure recommendation comes from a particular emphasis on the following:
1. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001;345:1368-77.
2. Gao F, Melody T, Daniels DF, Giles S, Fox S. The impact of compliance with 6-hour and 24-hour sepsis bundles on hospital mortality in patients with severe sepsis: a prospective observational study. Crit Care 2005;9:R764-70.
3. Sebat F, Johnson D, Musthafa AA, et al. A multidisciplinary community hospital program for early and rapid resuscitation of shock in nontrauma patients. Chest 2005;127:1729-43.
4. Kortgen A, Niederprum P, Bauer M. Implementation of an evidence-based "standard operating procedure" and outcome in septic shock. Crit Care Med 2006;34:943-9.
5. Shapiro NI, Howell MD, Talmor D, et al. Implementation and outcomes of the Multiple Urgent Sepsis Therapies (MUST) protocol. Crit Care Med 2006;34:1025-32.
6. Trzeciak S, Dellinger RP, Abate NL, et al. Translating research to clinical practice: a 1-year experience with implementing early goal-directed therapy for septic shock in the emergency department. Chest 2006;129:225-32.
7. Micek ST, Roubinian N, Heuring T, et al. Before-after study of a standardized hospital order set for the management of septic shock. Crit Care Med 2006;34:2707-13.
8. Lin SM, Huang CD, Lin HC, Liu CY, Wang CH, Kuo HP. A modified goal-directed protocol improves clinical outcomes in intensive care unit patients with septic shock: a randomized controlled trial. Shock 2006;26:551-7.
9. Qu HP, Qin S, Min D, Tang YQ. [The effects of earlier resuscitation on following therapeutic response in sepsis with hypoperfusion]. Zhonghua Wai Ke Za Zhi 2006;44:1193-6.
10. Nguyen HB, Corbett SW, Steele R, et al. Implementation of a bundle of quality indicators for the early management of severe sepsis and septic shock is associated with decreased mortality. Crit Care Med 2007;35:1105-12.
11. Chen ZQ, Jin YH, Chen H, Fu WJ, Yang H, Wang RT. [Early goal-directed therapy lowers the incidence, severity and mortality of multiple organ dysfunction syndrome]. Nan Fang Yi Ke Da Xue Xue Bao 2007;27:1892-5.
12. Jones AE, Focht A, Horton JM, Kline JA. Prospective external validation of the clinical effectiveness of an emergency department-based early goal-directed therapy protocol for severe sepsis and septic shock. Chest 2007;132:425-32.
13. Sebat F, Musthafa AA, Johnson D, et al. Effect of a rapid response system for patients in shock on time to treatment and mortality during 5 years. Crit Care Med 2007;35:2568-75.
14. El Solh AA, Akinnusi ME, Alsawalha LN, Pineda LA. Outcome of septic shock in older adults after implementation of the sepsis

"bundle". J Am Geriatr Soc 2008;56:272-8.
15. He ZY, Gao Y, Wang XR, Hang YN. [Clinical evaluation of execution of early goal directed therapy in septic shock]. Zhongguo Wei Zhong Bing Ji Jiu Yi Xue 2007;19:14-6.
16. Castro R, Regueira T, Aguirre ML, et al. An evidence-based resuscitation algorithm applied from the emergency room to the ICU improves survival of severe septic shock. Minerva Anestesiol 2008;74:223-31.
17. Zambon M, Ceola M, Almeida-de-Castro R, Gullo A, Vincent JL. Implementation of the Surviving Sepsis Campaign guidelines for severe sepsis and septic shock: we could go faster. J Crit Care 2008;23:455-60.
18. Zubrow MT, Sweeney TA, Fulda GJ, et al. Improving care of the sepsis patient. Jt Comm J Qual Patient Saf 2008;34:187-91.
19. Peel M. Care bundles: resuscitation of patients with severe sepsis. Nurs Stand 2008;23:41-6.
20. Focht A, Jones AE, Lowe TJ. Early goal-directed therapy: improving mortality and morbidity of sepsis in the emergency department. Jt Comm J Qual Patient Saf 2009;35:186-91.
21. Moore LJ, Jones SL, Kreiner LA, et al. Validation of a screening tool for the early identification of sepsis. J Trauma 2009;66:1539-46; discussion 46-7.
22. Puskarich MA, Marchick MR, Kline JA, Steuerwald MT, Jones AE. One year mortality of patients treated with an emergency department based early goal directed therapy protocol for severe sepsis and septic shock: a before and after study. Crit Care 2009;13:R167.
23. Ferrer R, Artigas A, Levy MM, et al. Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain. Jama 2008;299:2294-303.
24. Girardis M, Rinaldi L, Donno L, et al. Effects on management and outcome of severe sepsis and septic shock patients admitted to the intensive care unit after implementation of a sepsis program: a pilot study. Crit Care 2009;13:R143.
25. Wang JL, Chin CS, Chang MC, et al. Key process indicators of mortality in the implementation of protocol-driven therapy for severe sepsis. J Formos Med Assoc 2009;108:778-87.
26. Thiel SW, Asghar MF, Micek ST, Reichley RM, Doherty JA, Kollef MH. Hospital-wide impact of a standardized order set for the management of bacteremic severe sepsis. Crit Care Med 2009;37:819-24.
27. Pestana D, Espinosa E, Sanguesa-Molina JR, et al. Compliance With a Sepsis Bundle and Its Effect on Intensive Care Unit Mortality in Surgical Septic Shock Patients. J Trauma 2010.
28. Castellanos-Ortega A, Suberviola B, Garcia-Astudillo LA, et al. Impact of the Surviving Sepsis Campaign protocols on hospital length of stay and mortality in septic shock patients: results of a three-year follow-up quasi-experimental study. Crit Care Med 2010;38:1036-43.
29. Lefrant JY, Muller L, Raillard A, et al. Reduction of the severe sepsis or septic shock associated mortality by reinforcement of the recommendations bundle: A multicenter study. Ann Fr Anesth Reanim 2010.
30. Cardoso T, Carneiro AH, Ribeiro O, Teixeira-Pinto A, Costa-Pereira A. Reducing mortality in severe sepsis with the implementation of a core 6-hour bundle: results from the Portuguese community-acquired sepsis study (SACiUCI study). Crit Care 2010;14:R83.
31. [The effect of early goal-directed therapy on treatment of critical patients with severe sepsis/septic shock: a multi-center, prospective, randomized, controlled study]. Zhongguo Wei Zhong Bing Ji Jiu Yi Xue 2010;22:331-4.
32. Patel GW, Roderman N, Gehring H, Saad J, Bartek W. Assessing the Effect of the Surviving Sepsis Campaign Treatment Guidelines on Clinical Outcomes in a Community Hospital (November). Ann Pharmacother 2010.
33. Crowe CA, Mistry CD, Rzechula K, Kulstad CE. Evaluation of a modified early goal-directed therapy protocol. Am J Emerg Med 2010;28:689-93.
34. Daniels R, Nutbeam T, McNamara G, Galvin C. The sepsis six and the severe sepsis resuscitation bundle: a prospective observational cohort study. Emerg Med J 2011;28:507-12.
35. Gerber K. Surviving sepsis: a trust-wide approach. A multi-disciplinary team approach to implementing evidence-based guidelines. Nurs Crit Care 2010;15:141-51.
36. Gurnani PK, Patel GP, Crank CW, et al. Impact of the implementation of a sepsis protocol for the management of fluid-refractory septic shock: A single-center, before-and-after study. Clin Ther 2010;32:1285-93.
37. Levy MM, Dellinger RP, Townsend SR, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. Critical Care Medicine 2010;38:367-74.
38. Macredmond R, Hollohan K, Stenstrom R, Nebre R, Jaswal D, Dodek P. Introduction of a comprehensive management protocol for severe sepsis is associated with sustained improvements in timeliness of care and survival. Qual Saf Health Care 2010.
39. Mikkelsen ME, Gaieski DF, Goyal M, et al. Factors associated with nonadherence to early goal-directed therapy in the ED. Chest 2010;138:551-8.
40. Coba V, Whitmill M, Mooney R, et al. Resuscitation Bundle Compliance in Severe Sepsis and Septic Shock: Improves Survival, Is Better Late than Never. J Intensive Care Med 2011.
41. Sivayoham N, Rhodes A, Jaiganesh T, van Zyl Smit N, Elkhodhair S, Krishnanandan S. Outcomes from implementing early goal-directed therapy for severe sepsis and septic shock : a 4-year observational cohort study. Eur J Emerg Med 2011.

42. Westphal GA, Koenig A, Caldeira Filho M, et al. Reduced mortality after the implementation of a protocol for the early detection of severe sepsis. J Crit Care 2011;26:76-81.
43. Castellanos-Ortega A, Suberviola B, Garcia-Astudillo LA, Ortiz F, Llorca J, Delgado-Rodriguez M. Late compliance with the sepsis resuscitation bundle: impact on mortality. Shock 2011;36:542-7.
44. Jones AE, Troyer JL, Kline JA. Cost-effectiveness of an emergency department-based early sepsis resuscitation protocol. Crit Care Med 2011;39:1306-12.
45. O´Neill R, Morales J, Jule M. Early Goal-directed Therapy (EGDT) for Severe Sepsis/Septic Shock: Which Components of Treatment are More Difficult to Implement in a Community-based Emergency Department? J Emerg Med 2011.
46. Casserly B, Baram M, Walsh P, Sucov A, Ward NS, Levy MM. Implementing a collaborative protocol in a sepsis intervention program: lessons learned. Lung 2011;189:11-9.
47. Schramm GE, Kashyap R, Mullon JJ, Gajic O, Afessa B. Septic shock: A multidisciplinary response team and weekly feedback to clinicians improve the process of care and mortality. Crit Care Med 2011;39:252-8.
48. Suarez D, Ferrer R, Artigas A, et al. Cost-effectiveness of the Surviving Sepsis Campaign protocol for severe sepsis: a prospective nation-wide study in Spain. Intensive Care Med 2011;37:444-52.
49. Nguyen HB, Kuan WS, Batech M, et al. Outcome effectiveness of the severe sepsis resuscitation bundle with addition of lactate clearance as a bundle item: a multi-national evaluation. Crit Care 2011;15:R229.
50. Shiramizo SC, Marra AR, Durao MS, Paes AT, Edmond MB, Pavao dos Santos OF. Decreasing mortality in severe sepsis and septic shock patients by implementing a sepsis bundle in a hospital setting. PLoS ONE 2011;6:e26790.
51. Tromp M, Tjan DH, van Zanten AR, et al. The effects of implementation of the Surviving Sepsis Campaign in the Netherlands. Neth J Med 2011;69:292-8.
52. Winterbottom F, Seoane L, Sundell E, Niazi J, Nash T. Improving Sepsis Outcomes for Acutely Ill Adults Using Interdisciplinary Order Sets. Clin Nurse Spec 2011;25:180-5.
53. Bastani A, Galens S, Rocchini A, et al. ED identification of patients with severe sepsis/septic shock decreases mortality in a community hospital. Am J Emerg Med 2011.
54. Jeon K, Shin TG, Sim MS, et al. Improvements in Compliance of Resuscitation Bundles and Achievement of End Points After an Educational Program on the Management of Severe Sepsis and Septic Shock. Shock 2012.
55. Cannon CM, for the Multicenter Severe S, Septic Shock Collaborative G. The GENESIS Project (GENeralization of Early Sepsis InterventionS): A Multicenter Quality Improvement Collaborative. Acad Emerg Med 2010;17:1258.

1c.7 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* DETERMINATION OF QUALITY OF EVIDENCE
UNDERLYING METHODOLOGY:
A. RCT
B. Downgraded RCT or upgraded observational studies
C. Well-done observational studies
D. Case series or expert opinion

FACTORS THAT MAY DECREASE THE STREGNTH OF EVIDENCE:
1. Poor quality of planning and implementation of available RCTs, suggesting high likelihood of bias
2. Inconsistency of results (including problems with subgroup analyses)
3. Indirectness of evidence (differing population, intervention, control, outcomes, comparison)
4. Imprecision of results
5. High likelihood of reporting bias

MAIN FACTORS THAT MAY INCREASE STRENGTH OF EVIDENCE:
1. Large magnitude of effect (direct evidence, RR 2 with no plausible confounders)
2. Very large magnitude of effect with RR 5 and no threats to validity (by two levels)
3. Dose-response gradient RCT, randomized controlled trial; RR, relative risk.

FACTORS DETERMINING STRONG VS WEAK RECOMMENDATION:
1. Quality of evidence: the lower the quality of evidence, the less likely a strong recommendation
2. Relative importance of the outcomes: if values and preferences vary widely, a strong recommendation becomes less likely

3. Baseline risks of outcomes: the higher the risk, the greater the magnitude of benefit
4. Magnitude of relative risk, including benefits, harms, and burden: larger relative risk reductions or larger increases in relative risk of harm make a strong recommendation more or less likely, respectively
5. Absolute magnitude of the effect: the larger the absolute benefits and harms, the greater or lesser likelihood, respectively, of a strong recommendation
6. Precision of the estimates of the effects: the greater the precision, the more likely a strong recommendation
7. Costs: the higher the cost of treatment, the less likely a strong recommendation

1c.8 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* Although there is no explicit statement in the Surviving Sepsis Campaign (SSC)2008 guidelines regarding the overall consistency of results across studies supporting the guideline recommendations, the development of the SSC 2008 guidelines was by a committee of 68 international experts using the modified Delphi process in developing recommendations for the best current care of patients with severe sepsis and septic shock. These individuals represented 29 Sponsoring organizations: American Association of Critical-Care Nurses, American College of Chest Physicians, American College of Emergency Physicians, American Thoracic Society, Asia Pacific Association of Critical Care Medicine, Australian and New Zealand Intensive Care Societies, Brazilian Society of Critical Care, Canadian Critical Care Society, Chinese Society of Critical Care Medicine, Chinese Society of Critical Care Medicine - China Medical Association, Emirates Intensive Care Society, European Respiratory Society, European Society of Clinical Microbiology and Infectious Diseases, European Society of Intensive Care Medicine, European Society of Pediatric and Neonatal Intensive Care, Infectious Diseases Society of America, Indian Society of Critical Care Medicine, International Pan Arabian Critical Care Medicine Society, Japanese Association for Acute Medicine, Japanese Society of Intensive Care Medicine, Pediatric Acute Lung Injury and Sepsis Investigators, Society for Academic Emergency Medicine, Society of Critical Care Medicine, Society of Hospital Medicine, Surgical Infection Society, World Federation of Critical Care Nurses, World Federation of Pediatric Intensive and Critical Care Societies; World Federation of Societies of Intensive and Critical Care Medicine. Participation and endorsement by the German Sepsis Society and the Latin American Sepsis Institute.

Over the last decade the external validity and generalizability of the various components of the early management bundle have been established in over 50 publications containing over 20,000 patients in community and tertiary hospitals, ED and ICU settings, and medical and surgical patients. In addition, a meta-analysis found that in eight unblinded trials, one randomized and seven with historical controls, sepsis bundles were associated with a consistent (I2 = 0%, p = .87) and significant increase in survival (odds ratio, 1.91; 95% confidence interval, 1.49-2.45; p < .0001). For all studies reporting such data, there were consistent (I2 = 0%, p > or = .64) decreases in time to antibiotics, and increases in the appropriateness of antibiotics (p < or = .0002 for both).(2) Similar findings were noted in a meta-analysis by Chamberlain et al.(3)

1. Dellinger RP, Levy MM, Carlet JM, Bion J, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock. Crit Care Med. 2008 Jan;36(1):296-327.
2. Barochia AV, Cui X, Vitberg D, et al. Bundled care for septic shock: an analysis of clinical trials. Crit Care Med. Feb 2010;38(2):668-678.
3. Chamberlain DJ, Willis EM, Bersten AB. The severe sepsis bundles as processes of care: A meta-analysis. Aust Crit Care. Feb 14 2011.

1c.9 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)***:**
A meta-analysis found that in eight unblinded trials, one randomized and seven with historical controls, sepsis bundles were associated with a consistent (I2 = 0%, p = .87) and significant increase in survival (odds ratio, 1.91; 95% confidence interval, 1.49-2.45; p < .0001).(1) Similar findings were noted in a more recent and larger meta-analysis by Chamberlain.(7) In the presence of septic shock each hour delay in achieving administration of effective antibiotics is associated with a measurable 7.6% increase in mortality.(2) Although restriction of antibiotics as a strategy to reduce the development of antimicrobial resistance or to reduce cost is not an appropriate initial strategy in this patient population, once the causative pathogen has been identified, it may become apparent that none of the empirical drugs offers optimal therapy; that is, there may be another drug proven to produce superior clinical outcome that should therefore replace empirical agents. Narrowing the spectrum of antibiotic coverage and reducing the duration of antibiotic therapy will reduce the likelihood that the patient will develop superinfection with pathogenic or resistant organisms, such as Candida species, Clostridium difficile, or vancomycin-resistant Enterococcus faecium. However, the desire to minimize superinfections and other complications should not take precedence over the need to give the patient an adequate course of therapy to cure the infection that caused the severe sepsis or septic shock.3 After adjustment for baseline characteristics, administration of broad-spectrum antibiotics (OR, 0.86; 95%, CI 0.79– 0.93; p .0001), obtaining blood cultures before their initiation (OR, 0.76; 95% CI, 0.70 – 0.83; p < .0001) were all associated with lower hospital

mortality.(4) Blood pressure and lactate targets are predictors of outcome, detect early organ dysfunction and sudden hemodynamic compensation. Early aggressive fluid therapy is associated with improved outcomes over later aggressive fluid therapy.(5) ScvO2 is one of the most important bundle elements, predictive of outcome, and is superior to physical examination in detecting low cardiac index.(6)Patients attaining an ScvO2 of 70% have a two-fold improved mortality than patients treating without it.(7)

1. Barochia AV, Cui X, Vitberg D, et al. Bundled care for septic shock: an analysis of clinical trials. Crit Care Med. Feb 2010;38(2):668-678.
2. Kumar A, Roberts D, Wood KE, et al: Duration of hypotension prior to initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med 2006; 34:1589–1596.
3. Dellinger RP, Levy MM, Carlet JM, Bion J, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock. Crit Care Med. 2008 Jan;36(1):296-327.
4. Levy MM, Dellinger RP, Townsend SR, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. Critical Care Medicine 2010;38:367-74.
5. Murphy C, Schramm G, Doherty J, et al. The Importance of Fluid Management in Acute Lung Injury Secondary to Septic Shock. CHEST. 2009 July; vol 136, no. 1;102-109
6. Rivers EP, Katranji M, Jaehne KA, Brown S, Abou Dagher G, Cannon C, Coba V. Early interventions in severe sepsis and septic shock: a review of the evidence one decade later. Minerva Anestesiol. 2012 Jun;78(6):712-24
7. Chamberlain DJ, Willis EM, Bersten AB. The severe sepsis bundles as processes of care: A meta-analysis. Aust Crit Care. Feb 14 2011.

1c.10 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.11 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: The Surviving Sepsis Campaign is comprised of a consensus committee of over 50 international experts using the modified Delphi process. These individuals represented 29 Sponsoring organizations: American Association of Critical-Care Nurses, American College of Chest Physicians, American College of Emergency Physicians, American Thoracic Society, Asia Pacific Association of Critical Care Medicine, Australian and New Zealand Intensive Care Societies, Brazilian Society of Critical Care, Canadian Critical Care Society, Chinese Society of Critical Care Medicine, Chinese Society of Critical Care Medicine - China Medical Association, Emirates Intensive Care Society, European Respiratory Society, European Society of Clinical Microbiology and Infectious Diseases, European Society of Intensive Care Medicine, European Society of Pediatric and Neonatal Intensive Care, Infectious Diseases Society of America, Indian Society of Critical Care Medicine, International Pan Arabian Critical Care Medicine Society, Japanese Association for Acute Medicine, Japanese Society of Intensive Care Medicine, Pediatric Acute Lung Injury and Sepsis Investigators, Society for Academic Emergency Medicine, Society of Critical Care Medicine, Society of Hospital Medicine, Surgical Infection Society, World Federation of Critical Care Nurses, World Federation of Pediatric Intensive and Critical Care Societies; World Federation of Societies of Intensive and Critical Care Medicine. Participation and endorsement by the German Sepsis Society and the Latin American Sepsis Institute.

The 2008 guidelines process was funded fully by the Society of Critical Care Medicine. No industry funding was accepted or utilized. Nominal groups were assembled at key international meetings (for those committee members attending the conference). A stand-alone meeting was held for all sub-group heads, co- and vice chairs, and selected key individuals. Teleconferences and electronic-based discussion among subgroups and among the entire committee served as an integral part of the development. Methods: The Grading of Recommendations Assessment, Development and Evaluation. (GRADE) system to guide assessment of quality of evidence from high (A) to very low (D) and to determine the strength of recommendations was used. A strong recommendation (1) indicates that an intervention's desirable effects clearly outweigh its undesirable effects (risk, burden, cost) or clearly do not. Weak recommendations (2) indicate that the tradeoff between desirable and undesirable effects is less clear. Some recommendations are ungraded (UG). The grade of strong or weak is considered of greater clinical importance than a difference in letter level of quality of evidence. In areas without complete agreement, a formal process of resolution was developed and applied. Recommendations are in 3 groups: 1) those directly targeting severe sepsis; 2) recommendations targeting general care of the critically ill patient that are considered high priority in severe sepsis; and 3) pediatric considerations. A formal conflict of interest policy (COI) was developed at the onset of the process and enforced throughout. The entire guidelines process was conducted independent of any industry funding.

1c.12 System Used for Grading the Body of Evidence: GRADE

1c.13 If other, identify and describe the grading scale with definitions:

1c.14 Grade Assigned to the Body of Evidence: EARLY MANAGEMENT WITHIN 6 HOURS=1C, MEASURE LACTATE=1C, BLOOD CULTURES=1C, ANTIBIOTICS=1B, FLUIDS=1B, VASOPRESSORS=1D, MEASURE CVP & ScVO2=1C

1c.15 Summary of Controversy/Contradictory Evidence:

1c.16 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

1. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001;345:1368-77.
2. Gao F, Melody T, Daniels DF, Giles S, Fox S. The impact of compliance with 6-hour and 24-hour sepsis bundles on hospital mortality in patients with severe sepsis: a prospective observational study. Crit Care 2005;9:R764-70.
3. Sebat F, Johnson D, Musthafa AA, et al. A multidisciplinary community hospital program for early and rapid resuscitation of shock in nontrauma patients. Chest 2005;127:1729-43.
4. Kortgen A, Niederprum P, Bauer M. Implementation of an evidence-based "standard operating procedure" and outcome in septic shock. Crit Care Med 2006;34:943-9.
5. Shapiro NI, Howell MD, Talmor D, et al. Implementation and outcomes of the Multiple Urgent Sepsis Therapies (MUST) protocol. Crit Care Med 2006;34:1025-32.
6. Trzeciak S, Dellinger RP, Abate NL, et al. Translating research to clinical practice: a 1-year experience with implementing early goal-directed therapy for septic shock in the emergency department. Chest 2006;129:225-32.
7. Micek ST, Roubinian N, Heuring T, et al. Before-after study of a standardized hospital order set for the management of septic shock. Crit Care Med 2006;34:2707-13.
8. Lin SM, Huang CD, Lin HC, Liu CY, Wang CH, Kuo HP. A modified goal-directed protocol improves clinical outcomes in intensive care unit patients with septic shock: a randomized controlled trial. Shock 2006;26:551-7.
9. Qu HP, Qin S, Min D, Tang YQ. [The effects of earlier resuscitation on following therapeutic response in sepsis with hypoperfusion]. Zhonghua Wai Ke Za Zhi 2006;44:1193-6.
10. Nguyen HB, Corbett SW, Steele R, et al. Implementation of a bundle of quality indicators for the early management of severe sepsis and septic shock is associated with decreased mortality. Crit Care Med 2007;35:1105-12.
11. Chen ZQ, Jin YH, Chen H, Fu WJ, Yang H, Wang RT. [Early goal-directed therapy lowers the incidence, severity and mortality of multiple organ dysfunction syndrome]. Nan Fang Yi Ke Da Xue Xue Bao 2007;27:1892-5.
12. Jones AE, Focht A, Horton JM, Kline JA. Prospective external validation of the clinical effectiveness of an emergency department-based early goal-directed therapy protocol for severe sepsis and septic shock. Chest 2007;132:425-32.
13. Sebat F, Musthafa AA, Johnson D, et al. Effect of a rapid response system for patients in shock on time to treatment and mortality during 5 years. Crit Care Med 2007;35:2568-75.
14. El Solh AA, Akinnusi ME, Alsawalha LN, Pineda LA. Outcome of septic shock in older adults after implementation of the sepsis "bundle". J Am Geriatr Soc 2008;56:272-8.
15. He ZY, Gao Y, Wang XR, Hang YN. [Clinical evaluation of execution of early goal directed therapy in septic shock]. Zhongguo Wei Zhong Bing Ji Jiu Yi Xue 2007;19:14-6.
16. Castro R, Regueira T, Aguirre ML, et al. An evidence-based resuscitation algorithm applied from the emergency room to the ICU improves survival of severe septic shock. Minerva Anestesiol 2008;74:223-31.
17. Zambon M, Ceola M, Almeida-de-Castro R, Gullo A, Vincent JL. Implementation of the Surviving Sepsis Campaign guidelines for severe sepsis and septic shock: we could go faster. J Crit Care 2008;23:455-60.
18. Zubrow MT, Sweeney TA, Fulda GJ, et al. Improving care of the sepsis patient. Jt Comm J Qual Patient Saf 2008;34:187-91.
19. Peel M. Care bundles: resuscitation of patients with severe sepsis. Nurs Stand 2008;23:41-6.
20. Focht A, Jones AE, Lowe TJ. Early goal-directed therapy: improving mortality and morbidity of sepsis in the emergency department. Jt Comm J Qual Patient Saf 2009;35:186-91.
21. Moore LJ, Jones SL, Kreiner LA, et al. Validation of a screening tool for the early identification of sepsis. J Trauma 2009;66:1539-46; discussion 46-7.
22. Puskarich MA, Marchick MR, Kline JA, Steuerwald MT, Jones AE. One year mortality of patients treated with an emergency department based early goal directed therapy protocol for severe sepsis and septic shock: a before and after study. Crit Care 2009;13:R167.
23. Ferrer R, Artigas A, Levy MM, et al. Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain. Jama 2008;299:2294-303.
24. Girardis M, Rinaldi L, Donno L, et al. Effects on management and outcome of severe sepsis and septic shock patients admitted to the intensive care unit after implementation of a sepsis program: a pilot study. Crit Care 2009;13:R143.
25. Wang JL, Chin CS, Chang MC, et al. Key process indicators of mortality in the implementation of protocol-driven therapy for

severe sepsis. J Formos Med Assoc 2009;108:778-87.
26. Thiel SW, Asghar MF, Micek ST, Reichley RM, Doherty JA, Kollef MH. Hospital-wide impact of a standardized order set for the management of bacteremic severe sepsis. Crit Care Med 2009;37:819-24.
27. Pestana D, Espinosa E, Sanguesa-Molina JR, et al. Compliance With a Sepsis Bundle and Its Effect on Intensive Care Unit Mortality in Surgical Septic Shock Patients. J Trauma 2010.
28. Castellanos-Ortega A, Suberviola B, Garcia-Astudillo LA, et al. Impact of the Surviving Sepsis Campaign protocols on hospital length of stay and mortality in septic shock patients: results of a three-year follow-up quasi-experimental study. Crit Care Med 2010;38:1036-43.
29. Lefrant JY, Muller L, Raillard A, et al. Reduction of the severe sepsis or septic shock associated mortality by reinforcement of the recommendations bundle: A multicenter study. Ann Fr Anesth Reanim 2010.
30. Cardoso T, Carneiro AH, Ribeiro O, Teixeira-Pinto A, Costa-Pereira A. Reducing mortality in severe sepsis with the implementation of a core 6-hour bundle: results from the Portuguese community-acquired sepsis study (SACiUCI study). Crit Care 2010;14:R83.
31. [The effect of early goal-directed therapy on treatment of critical patients with severe sepsis/septic shock: a multi-center, prospective, randomized, controlled study]. Zhongguo Wei Zhong Bing Ji Jiu Yi Xue 2010;22:331-4.
32. Patel GW, Roderman N, Gehring H, Saad J, Bartek W. Assessing the Effect of the Surviving Sepsis Campaign Treatment Guidelines on Clinical Outcomes in a Community Hospital (November). Ann Pharmacother 2010.
33. Crowe CA, Mistry CD, Rzechula K, Kulstad CE. Evaluation of a modified early goal-directed therapy protocol. Am J Emerg Med 2010;28:689-93.
34. Daniels R, Nutbeam T, McNamara G, Galvin C. The sepsis six and the severe sepsis resuscitation bundle: a prospective observational cohort study. Emerg Med J 2011;28:507-12.
35. Gerber K. Surviving sepsis: a trust-wide approach. A multi-disciplinary team approach to implementing evidence-based guidelines. Nurs Crit Care 2010;15:141-51.
36. Gurnani PK, Patel GP, Crank CW, et al. Impact of the implementation of a sepsis protocol for the management of fluid-refractory septic shock: A single-center, before-and-after study. Clin Ther 2010;32:1285-93.
37. Levy MM, Dellinger RP, Townsend SR, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. Critical Care Medicine 2010;38:367-74.
38. Macredmond R, Hollohan K, Stenstrom R, Nebre R, Jaswal D, Dodek P. Introduction of a comprehensive management protocol for severe sepsis is associated with sustained improvements in timeliness of care and survival. Qual Saf Health Care 2010.
39. Mikkelsen ME, Gaieski DF, Goyal M, et al. Factors associated with nonadherence to early goal-directed therapy in the ED. Chest 2010;138:551-8.
40. Coba V, Whitmill M, Mooney R, et al. Resuscitation Bundle Compliance in Severe Sepsis and Septic Shock: Improves Survival, Is Better Late than Never. J Intensive Care Med 2011.
41. Sivayoham N, Rhodes A, Jaiganesh T, van Zyl Smit N, Elkhodhair S, Krishnanandan S. Outcomes from implementing early goal-directed therapy for severe sepsis and septic shock : a 4-year observational cohort study. Eur J Emerg Med 2011.
42. Westphal GA, Koenig A, Caldeira Filho M, et al. Reduced mortality after the implementation of a protocol for the early detection of severe sepsis. J Crit Care 2011;26:76-81.
43. Castellanos-Ortega A, Suberviola B, Garcia-Astudillo LA, Ortiz F, Llorca J, Delgado-Rodriguez M. Late compliance with the sepsis resuscitation bundle: impact on mortality. Shock 2011;36:542-7.
44. Jones AE, Troyer JL, Kline JA. Cost-effectiveness of an emergency department-based early sepsis resuscitation protocol. Crit Care Med 2011;39:1306-12.
45. O´Neill R, Morales J, Jule M. Early Goal-directed Therapy (EGDT) for Severe Sepsis/Septic Shock: Which Components of Treatment are More Difficult to Implement in a Community-based Emergency Department? J Emerg Med 2011.
46. Casserly B, Baram M, Walsh P, Sucov A, Ward NS, Levy MM. Implementing a collaborative protocol in a sepsis intervention program: lessons learned. Lung 2011;189:11-9.
47. Schramm GE, Kashyap R, Mullon JJ, Gajic O, Afessa B. Septic shock: A multidisciplinary response team and weekly feedback to clinicians improve the process of care and mortality. Crit Care Med 2011;39:252-8.
48. Suarez D, Ferrer R, Artigas A, et al. Cost-effectiveness of the Surviving Sepsis Campaign protocol for severe sepsis: a prospective nation-wide study in Spain. Intensive Care Med 2011;37:444-52.
49. Nguyen HB, Kuan WS, Batech M, et al. Outcome effectiveness of the severe sepsis resuscitation bundle with addition of lactate clearance as a bundle item: a multi-national evaluation. Crit Care 2011;15:R229.
50. Shiramizo SC, Marra AR, Durao MS, Paes AT, Edmond MB, Pavao dos Santos OF. Decreasing mortality in severe sepsis and septic shock patients by implementing a sepsis bundle in a hospital setting. PLoS ONE 2011;6:e26790.
51. Tromp M, Tjan DH, van Zanten AR, et al. The effects of implementation of the Surviving Sepsis Campaign in the Netherlands.

Neth J Med 2011;69:292-8.
52. Winterbottom F, Seoane L, Sundell E, Niazi J, Nash T. Improving Sepsis Outcomes for Acutely Ill Adults Using Interdisciplinary Order Sets. Clin Nurse Spec 2011;25:180-5.
53. Bastani A, Galens S, Rocchini A, et al. ED identification of patients with severe sepsis/septic shock decreases mortality in a community hospital. Am J Emerg Med 2011.
54. Jeon K, Shin TG, Sim MS, et al. Improvements in Compliance of Resuscitation Bundles and Achievement of End Points After an Educational Program on the Management of Severe Sepsis and Septic Shock. Shock 2012.
55. Cannon CM, for the Multicenter Severe S, Septic Shock Collaborative G. The GENESIS Project (GENeralization of Early Sepsis InterventionS): A Multicenter Quality Improvement Collaborative. Acad Emerg Med 2010;17:1258.

1c.17 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)*:

GOALS OF INITIAL RESUSCITATION (Strong Recommendation):
This protocol should be initiated as soon as hypoperfusion is recognized and should not be delayed pending ICU admission. During the first 6 hrs of resuscitation, the goals of initial resuscitation of sepsis-induced hypoperfusion should include all of the following as one part of a treatment protocol: central venous pressure 8–12 mm Hg, mean arterial pressure (MAP) >=65 mm Hg, urine output >=0.5 mL·kg-1·hr-1, central venous (superior vena cava) or mixed venous oxygen saturation >=70% or >=65%, respectively (grade 1C).

MEASURE LACTATE (Strong Recommendation):
We recommend the protocolized resuscitation of a patient with sepsis induced shock, defined as tissue hypoperfusion (hypotension persisting after initial fluid challenge or blood lactate concentration > or =4 mmol/L). (grade 1C).

APPROPRIATE BLOOD CULTURES (Strong Recommendation):
We recommend obtaining appropriate cultures before antimicrobial therapy is initiated if such cultures do not cause significant delay in antibiotic administration. To optimize identification of causative organisms, we recommend at least two blood cultures be obtained before antibiotics with at least one drawn percutaneously and one drawn through each vascular access device, unless the device was recently (<48 hrs) inserted. Cultures of other sites (preferably quantitative where appropriate), such as urine, cerebrospinal fluid, wounds, respiratory secretions, or other body fluids that may be the source of infection should also be obtained before antibiotic therapy if not associated with significant delay in antibiotic administration (grade 1C).

ANTIBIOTIC THERAPY (Strong Recommendations):
1. We recommend that intravenous antibiotic therapy be started as early as possible and within the first hour of recognition of septic shock (1B) and severe sepsis without septic shock (1D). Appropriate cultures should be obtained before initiating antibiotic therapy but should not prevent prompt administration of antimicrobial therapy (grade 1D).
2a. We recommend that initial empirical anti-infective therapy include one or more drugs that have activity against all likely pathogens (bacterial and/or fungal) and that penetrate in adequate concentrations into the presumed source of sepsis (grade 1B).

FLUID THERAPY (Strong Recommendation):
1. We recommend fluid resuscitation with either natural/artificial colloids or crystalloids. There is no evidence-based support for one type of fluid over another (grade 1B).
2. We recommend that fluid resuscitation initially target a central venous pressure of >=8 mm Hg (12 mm Hg in mechanically ventilated patients). Further fluid therapy is often required (grade 1C).
3a. We recommend that a fluid challenge technique be applied wherein fluid administration is continued as long as the hemodynamic improvement (e.g., arterial pressure, heart rate, urine output) continues (grade 1D).
3b. We recommend that fluid challenge in patients with suspected hypovolemia be started with >=1000 mL of crystalloids or 300–500 mL of colloids over 30 mins. More rapid administration and greater amounts of fluid may be needed in patients with sepsis-induced tissue hypoperfusion (see Initial Resuscitation recommendations)(grade 1D).
3c. We recommend that the rate of fluid administration be reduced substantially when cardiac filling pressures (central venous pressure or pulmonary artery balloon-occluded pressure) increase without concurrent hemodynamic improvement (grade 1D).

VASOPRESSORS (Strong Recommendations):
1. We recommend that mean arterial pressure (MAP) be maintained >65 mm Hg(grade 1C).
2. We recommend either norepinephrine or dopamine as the first choice vasopressor agent to correct hypotension in septic shock (administered through a central catheter as soon as one is available) (grade 1C).
5. We recommend that low-dose dopamine not be used for renal protection(grade 1A).
6. We recommend that all patients requiring vasopressors have an arterial catheter placed as soon as practical if resources are available

(grade 1D).

MEASURE CVP & MEASURE ScvO2: (Strong Recommendation)
1.We suggest that during the first 6 hrs of resuscitation of severe sepsis or septic shock, if ScvO2 or SCvO2 of 70% or 65%, respectively, is not achieved with fluid resuscitation to the central venous pressure target (1C)
2.If the ScvO2 is <70%, CVP and MAP goals are met; then transfusion of packed red blood cells to achieve a hematocrit of >=30% and/or administration of a dobutamine infusion (up to a maximum of 20 micrograms·kg-1·min-1) be used to achieve this goal (2C).

1c.18 Clinical Practice Guideline Citation: Dellinger RP, Levy MM, Carlet JM, Bion J, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock. Crit Care Med. 2008 Jan;36(1):296-327.

1c.19 National Guideline Clearinghouse or other URL: http://www.guideline.gov/content.aspx?id=12231&search=surviving+sepsis

1c.20 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.21 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: APPENDIX I: 2008 Surviving Ssepsis Campaign (SSC) Guidelines Committee:
R. Phillip Dellinger (Chair), Tom Ahrens(a), Naoki Aikawa(b), Derek Angus, Djillali Annane, Richard Beale, Gordon R. Bernard, Julian Bion(c), Christian Brun-Buisson, Thierry Calandra, Joseph Carcillo, Jean Carlet, Terry Clemmer, Jonathan Cohen, Edwin A. Deitch(d),Jean-Francois Dhainaut, Mitchell Fink, Satoshi Gando (b), Herwig Gerlach, Gordon Guyatt (e), Maurene Harvey, Jan Hazelzet, Hiroyuki Hirasawa,f Steven M. Hollenberg, Michael Howell, Roman Jaeschke (e), Robert Kacmarek, Didier Keh, Mitchell M. Levy (g), Jeffrey Lipman, John J. Marini, John Marshall, Claude Martin, Henry Masur, Steven Opal, Tiffany M. Osborn (h), Giuseppe Pagliarello (i), Margaret Parker, Joseph Parrillo, Graham Ramsay, Adrienne Randolph, Marco Ranieri, Robert C. Read (j), Konrad Reinhart (k), Andrew Rhodes, Emanuel Rivers (h), Gordon Rubenfeld, Jonathan Sevransky, Eliezer Silva,l Charles L. Sprung, B. Taylor Thompson, Sean R. Townsend, Jeffery Vender (m), Jean-Louis Vincent (n), Tobias Welte (o), Janice Zimmerman.

a American Association of Critical-Care Nurses;
b Japanese Association for Acute Medicine;
c European Society of Intensive Care Medicine;
d Surgical Infection Society;
e Grades of Recommendation, Assessment, Development and Evaluation (GRADE)
f Japanese Society of Intensive Care Medicine;
g Society of Critical Care Medicine;
h American College of Emergency Physicians;
i Canadian Critical Care Society;
j European Society of Clinical Microbiology and Infectious Diseases;
k German Sepsis Society;
l Latin American Sepsis Institute;
m American College of Chest Physicians;
n International Sepsis Forum;
o European Respiratory Society.

APPENDIX J: Author Disclosure Information 2006–2007
Dr. Dellinger has consulted for Astra-Zeneca, Talecris, and B Braun. He has received honoraria from Eli Lilly (2),Brahms (2), INO Therapeutics (1), Pulsion (1), and bioMerieux (1). He has also received grant support from AstraZeneca and Artisan. Dr. Levy has received honoraria from Eli Lilly and Edwards Lifesciences. He has also received grant support from Philips Medical Systems, Edwards Lifesciences, Philips Medical Systems, Novartis, Biosite, and Eisai.
Dr. Carlet has consulted for Forrest, Wyeth, Chiron, bioMerieux, and Glaxo-
SmithKline. He has also received honoraria from Eli Lilly, Becton Dickinson,
Jansen, Cook, AstraZeneca, Hutchinson, Bayer, Gilead, MSD, and Targanta. Dr. Bion has not disclosed any potential conflicts of interest. Dr. Parker has consulted for Johnson & Johnson. Dr. Jaeschke has received honoraria from AstraZeneca, Boehringer, Eli Lilly, GlaxoSmithKline, and MSD. Dr. Reinhart has consulted for Eli Lilly and Edwards Lifesciences. He has also received honoraria from B Braun and royalties from Edwards Lifesciences. Dr. Angus has consulted for or received speaking fees from AstraZeneca, BrahmsDiagnostica, Eisai, Eli Lilly, Glaxo-SmithKline, OrthoBiotech, Takeda, and Wyeth-Ayerst. He has also received grant support from

GlaxoSmithKline, Ortho-Biotech, and Amgen.
Dr. Brun-Buisson has not disclosed any potential conflicts of interest. Dr. Beale has received honoraria from Eisai and speaking fees (paid to university)
from Lilly UK, Philips, Lidco, and Chiron. Dr. Calandra has consulted for Baxter, received honoraria from Roche Diagnostics, and received grant support from Baxter and Roche Diagnostics. He also served on the advisory board for Biosite. Dr. Dhainaut has consulted for Eli Lilly and Novartis. He has also received honoraria from Eli Lilly. Dr. Gerlach has not disclosed any potential
conflicts of interest. Ms. Harvey has not disclosed any potential conflicts of interest. Dr. Marini has consulted for KCI and received honoraria from Maquet.
Dr. Marshall has consulted for Becton Dickinson, Takeda, Pfizer, Spectral Diagnostics, Eisai, and Leo-Pharma. He has also received honoraria from Spectral Diagnostics. Dr. Ranieri has served on the advisor board for Maquet and received support for a sponsored trial from Eli Lilly. He has also received grant support from Tyco, Draeger, and Hamilton. Dr. Ramsay has consulted for Edwards Lifesciences and Respironics. Dr. Sevransky has not disclosed any
potential conflicts of interest. Dr. Thompson has consulted for Eli Lilly, Abbott, and AstraZeneca. He has also received grant support from the NIH
for a study on computerized glucose control. Dr. Townsend has not disclosed any
potential conflicts of interest. Dr. Vender has consulted and received
honoraria from Eli Lilly. Dr. Zimmerman has not disclosed any potential conflicts of interest. Dr. Vincent has consulted for Astra-Zeneca, Biosite, bioMerieux, Edwards Lifesciences, Eli Lilly, Eisai, Ferring, Glaxo-SmithKline, Intercell, Merck, Novartis, NovoNordisk, Organon, Pfizer, Philips Medical Systems, Roche Diagnostics, Spectral Diagnostics, Takeda, and Wyeth-Lederle. He has also received honoraria from Eli Lilly, Edwards Lifesciences, Eisai, GlaxoSmithKline, Novartis, NovoNordisk, and Pfizer.

1c.22 System Used for Grading the Strength of Guideline Recommendation: GRADE

1c.23 If other, identify and describe the grading scale with definitions:

1c.24 Grade Assigned to the Recommendation: EARLY MANAGEMENT WITHIN 6 HOURS=1C, MEASURE LACTATE=1C, BLOOD CULTURES=1C, ANTIBIOTICS=1B, FLUIDS=1B, VASOPRESSORS=1D, MEASURE CVP & ScVO2=1C

1c.25 Rationale for Using this Guideline Over Others: It is the Henry Ford Hospital policy to use guidelines, which are evidence-based, actionable by facilities and health-care providers, and developed by a national specialty organization or government agency. In addition, the HFH also accepts as the evidence base for measures to include documented quality improvement (QI) initiatives or implementation projects that have demonstrated improvement in quality of care. There was strong agreement among a large cohort of 55 international experts regarding many recommendations for the best current care of patients with severe sepsis and septic shock. Evidence-based recommendations are the first step toward improved outcomes for this important group of critically ill patients.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?
1c.26 Quantity: High **1c.27** Quality: High **1c.28** Consistency: High

Was the threshold criterion, *Importance to Measure and Report*, met?
(*1a & 1b must be rated moderate or high and 1c yes*) **Yes☐ No☐**
Provide rationale based on specific subcriteria:

For a new measure if the Committee votes NO, then STOP.
For a measure undergoing endorsement maintenance, if the Committee votes NO because of 1b. (no opportunity for improvement), it may be considered for continued endorsement and all criteria need to be evaluated.

THIS FORM HAS BEEN UPDATED SINCE DATE OF SUBMISSION

UPDATED EVIDENCE

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** NQF #0500

Measure Title: Severe Sepsis and Septic Shock: Management Bundle
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:
Click here to enter composite measure #/ title
Date of Submission: 2/3/2016

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence sub criterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome

☐ Health outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

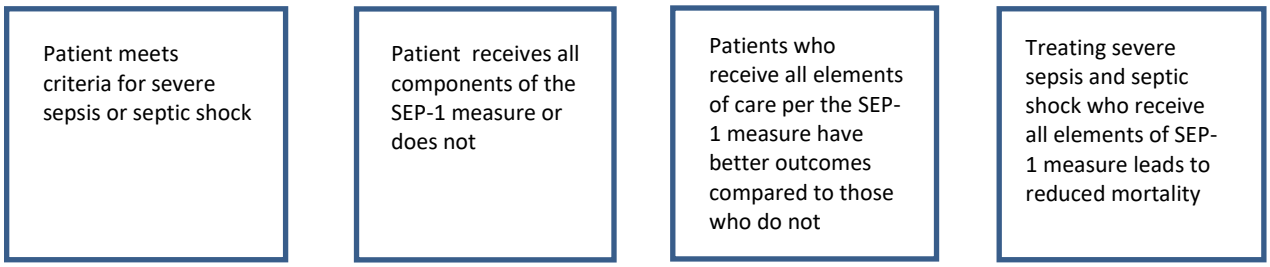
☒ Process: This measure focuses on adults 18 years and older with a diagnosis of severe sepsis or septic shock. Consistent with Surviving Sepsis Campaign guidelines, it assesses measurement of lactate, obtaining blood cultures, administering broad spectrum antibiotics, fluid resuscitation, vasopressor administration, reassessment of volume status and tissue perfusion, and repeat lactate measurement. As reflected in the data elements and their definitions, the first three interventions should occur within 3 hours of presentation of severe sepsis, while the remaining interventions are expected to occur within 6 hours of presentation of septic shock

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.



****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

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).

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	

Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

Please refer to the following section for a clinical rationale based on the surviving sepsis guidelines and bolstered by evidence from the literature which supports each of the unique elements which make up the SEP-1 measure as well as outcome studies which show the positive impact, of the bundles of care (part of SEP-1), on mortality.
Finally, please refer to the end for an independent analysis of trends in mortality based on data submitted to the Medicare Chronic Disease Warehouse (CDW), stratified by SEP-1 measure adherence.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

Evidence by Data Element Level-Updated From Last Submission

Measure lactate level

Measuring lactate levels in sepsis provides diagnostic, prognostic and therapeutic utility.[1] Lactate levels improve the risk stratification power of the systemic inflammatory response syndrome (SIRS) criteria.[2] Early risk stratification using lactate levels for undetected, hidden and untreated cardiovascular insufficiency (cryptic shock) is an important aspect of early sepsis care.[3] Early risk stratification helps detect these high risk patients especially in the hands of an inexperienced healthcare provider, or when the disease is undifferentiated.[4,5] Cryptic shock, which can occur in up to 20% of sepsis patients, leads to increased episodes of sudden respiratory and cardiovascular events. These patients are inadvertently sent to general practice floors only to suffer from acute hemodynamic deterioration.[6-8] These complications are the most preventable causes of death in the first 24 hours of sepsis care.[6,8-10] These complications are reduced by 50% or less, when early risk stratification is combined with appropriate intervention.[7,8] Whether central venous or arterial, increased lactate levels both initially and over time are associated with increased mortality.[11-17] Weil and Aduen et al. established the prognostic value of a lactate greater than or equal to 4 mM/Liter on hospital admission, which has been confirmed by

multiple follow up studies.[18-25] It has also be shown that intermediate lactate levels (2-4 mM/L) are also associated with increased mortality which is significantly reduced (19% odds ratio for hospital mortality) with protocolized care.[21,26-29] These patients are particularly at risk for mortality, as 22.7% will progress to sepsis-induced tissue hypoperfusion with an associated mortality of 10.1% . Lactate levels also reduce the time to antibiotics, time to fluid therapy and hemodynamic optimization. As a result, it has been associated with decreases in resource utilization and cost. [1,30,31] In follow-up unblinded clinical studies following the original EGDT study, early risk stratification with lactate was a requisite for enrollment in all treatment groups. There is significant mortality reduction simply by measuring a lactate level.[15,31-33] When this has been included as usual care as in recent sepsis trials, historically low mortalities were seen in all treatment groups.[34,35]

Obtain cultures prior to antibiotics

Appropriate routine microbiologic cultures (including blood) should be obtained before starting antimicrobial therapy if it results in no substantial delay in the start of antimicrobials.[36-41] This is because rapid sterilization of cultures can occur within minutes to hours after the first dose of an appropriate antimicrobial. [42,43] This must be balanced against the mortality risk of delaying a key therapy.[44,45] When antibiotics are used among septic emergency department (ED) patients, drug-resistant bacteria are covered infrequently.[46] Bacteremia is associated with increased mortality which may be increased up to five-fold in patients who receive inappropriate initial antibiotic therapy.[1,47-50] This is particularly important in candidemic patients.[51] In patients admitted to the ICU for sepsis, the adequacy of initial empirical antimicrobial treatment is crucial in terms of outcome.[52] The benefits also result from reducing complications associated with antibiotic use such as drug reactions, allergies, development of drug-resistant organisms and the occurrence of Clostridium difficile colitis.[52] Collecting blood cultures has been associated with improved outcomes because pathogens identified allow for customized therapy.[49,53,54] In addition, they enhance de-escalation which has been associated with improved survival.[55] In the first quarter of a multicenter quality improvement program for sepsis care, only 64.5% of patients had blood cultures collected.[1] The performance measure for collecting blood cultures for suspected sepsis has been previously used as a core component of multicenter and national quality improvement initiatives.[1,56]

Administer broad spectrum antibiotics

The evidence for early and appropriate antibiotic administration is abundantly present in both animal and multiple human studies of sepsis. Animal models have demonstrated that antibiotics alone and cardiovascular support alone are relatively ineffective in the treatment of septic shock. When combined, however, these two therapies provide moderately successful treatment for this highly lethal disorder.[57] Multiple observational studies of septic shock reveal a significant association between time to appropriate antibiotics, mortality and health care resource consumption.[1,44,49,56-62] Mortality can increase up to 7.6% for each hour delay in antibiotic administration after the on-set of hypotension or shock.[56] Timely antibiotic administration, whether in the ED or ICU, is not only associated with decreased mortality, but also hospital length of stay and costs.[49,52,56,59-66] In-spite of the existing knowledge, many patients with sepsis do not receive antibiotics until they arrive on the inpatient unit, and frequently with inadequate coverage.[46,67] Multicenter quality improvement projects reveal that timely administration of broad-spectrum antibiotics was associated with significantly higher risk adjusted survival.[1] Based on a preponderance of data, the current recommendation in the international guidelines for the management of severe sepsis and septic shock includes the administration of broad-spectrum antibiotic therapy within 1 h of diagnosis of septic shock and severe sepsis.[68,69]

Administer 30 ml/kg crystalloid for hypotension or lactate ≥ 4 mmol/L

A hypotensive episode or lactate greater than 4 mM/L is associated with an increased risk of death, and the response to an adequate fluid challenge improves upon this discriminatory value.[26,70-72] Early aggressive fluid therapy must be distinguished from late aggressive fluid therapy.[9,70,73-75] Multiple studies have shown that a prompt fluid challenge (30 cc/kg) is associated with increased MAP, normalization of $ScvO_2$ decreased vasopressor use at 6 hours and decreased need for dialysis.[70,76] This is also associated with an absolute mortality reduction of up to 1.4-6.2%, or a 15-31% relative reduction in hospital/30-day mortality and hospital length of stay.[26,70,72] These findings were seen even in patients with a history of renal, heart failure and acute lung injury.[76-79] The benefits of fluid administration are maximal when initiated within 30 minutes, making the fluid challenge an important aspect of early sepsis care. As a result, Lee et al. concluded that "earlier fluid resuscitation may account for the lack of outcome differences in the trio of EGDT trials and may have

contributed to the overall low 60-day in-hospital mortality rate of 19%".[70] In the first quarter of a multicenter quality improvement program for sepsis care, only 59.8% of patients received fluid resuscitation consistent with guidelines.[1] This performance measure has been previously used as a core component of multicenter and national quality improvement initiatives.[80]

Remeasure lactate if initial lactate is elevated

Previous seminal investigations have shown that lactate clearance in the ICU setting reflects the microcirculation and predicts outcome.[181-183] Other investigators have shown that early clearance of lactate over the first 6 hours after presentation is associated with a significant decrease in pro- and anti-inflammatory biomarkers, improved organ function and reduced mortality.[182,184,185] Multiple studies have confirmed these findings.[13,25,186-189] The use of lactate clearance as a goal to guide early therapy is associated with a reduction in the risk of death in adult patients with sepsis.[34,190]

Apply vasopressors (for hypotension that does not respond to initial fluid resuscitation to maintain a mean arterial pressure ≥ 65 mmHg)

Vasodilatation and the loss of systemic autoregulation necessitates exogenous administration of vasopressors. The duration of hypotension is directly associated with mortality.[81,82] While increasing MAP above 65 mmHg has been associated with increased cardiac output, improved microvascular function and decreased blood lactate concentrations, there is individual variation.[83] As a result, a mean arterial blood pressure target of 65 mmHg is the recommended target.[81,84-87] However, a MAP of around 75 to 85 mm Hg may reduce the development of acute kidney injury in patients with chronic arterial hypertension.[88] If the clinician chooses to provide additional fluids to resolve hypotension rather than vasopressors, this will be considered adequate for the measure.

Reassess volume status and tissue perfusion in the event of persistent hypotension (MAP < 65 mm Hg) after initial fluid administration or if initial lactate was ≥ 4 mmol/L

Recent trials of early protocolized care in sepsis have shown similar mortality reductions with alternative methods of volume and perfusion assessments in addition to $ScvO_2$ and CVP. As a result this measure will incorporate the options that reflect usual care or a breath of standards of practice (from community to academic tertiary care hospitals).[69,89]
The measure no longer requires a provider to state the method of reassessment used. Reassessment itself, regardless of method, is a best practice recommendation.[69,89] The clinician can attest that volume and perfusion reassessment has occurred, even without reference to the method used. This will meet the measure's volume and perfusion reassessment requirement. Some common practices used to reassess volume status and tissue perfusion (none are required) include:

- **Bedside Cardiovascular Ultrasound**

  Point of care ultrasound (POCUS), whether trans-thoracic or esophageal, is a non-invasive means of hemodynamic assessment.[90-95] Using a goal-directed approach by looking at the pump, tank and pipes (the RUSH exam), one can assess preload, after load and contractility.[92,96] Clinical management involving the early use of ultrasound in patients with sepsis significantly reduces physicians' diagnostic uncertainty improves procedural accuracy and substantially changes management and resource utilization.[97,98] Intra-cardiac, vena cava diameters, internal jugular vein, LVEDV area measurements after a fluid challenge, passive leg rising and respiratory variation may be used to asses volume status.[99-104] Left ventricular strain seen on cardiac ultrasound during sepsis is associated with a decreased $ScvO_2$ and increased lactate, rendering their assessment a potential impact on therapy.[97,105-107] Ultrasound is useful to assist in procedures, cardiac output (CO) measurement, detect myocardial dysfunction, pericardial disease, aortic disease, intra-peritoneal fluid and a pneumothorax.[94,108-111] POCUS is also an effective and reliable tool for the identification of septic source, and it is superior to the initial clinical evaluation alone.[107,112-115] For these reasons, POCUS has become part of the standard curriculum in residency and fellowship training.[90,116] However, there are not universal standards of competency, and expertise will vary by practitioner.[117]
- **Dynamic Assessment of Fluid Responsiveness**

  Pulse pressure variation (PPV) or stroke volume variation (SVV) during a positive pressure breath in the intubated patient can be used to predict the responsiveness of CO to changes in preload.[118] SVV is defined as the difference between the maximal pulse pressure and the minimum pulse pressure divided by the average of these two pressures.[118] In ventilated patients, measures of SVV using arterial pulse contour analysis estimates CO and can

demonstrate fluid responsiveness. [100] A SVV of 13% is highly sensitive and specific for detecting preload responsiveness.[119] SVV has been compared to CVP, PAOP, and systolic pressure variation as predictors of preload responsiveness. Patients are classified as preload responsive if their cardiac index increased by at least 10-15% after rapid infusion of a standard volume of intravenous fluid or passive leg rising.[118] Atrial arrhythmias and spontaneous breathing can interfere with the usefulness of this technique.[100]

- **Focused exam including vital signs, cardiopulmonary, capillary refill and skin findings**

Vital signs which include the shock index (SI; heart rate/systolic blood pressure has been evaluated for the identification of patients with septic shock in the ED and as a possible end point for resuscitation.[120] An SI of ≥ 0.7 has been shown to perform as well as systemic inflammatory response syndrome criteria in the ED and may be predictive of outcomes at values ≥ 1.0.[121] Rady showed that even after initial resuscitation with the appearance of normal vital signs, significant global tissue hypoxia would persist with an elevated SI.[122-124] ED patients with severe sepsis and a sustained SI elevation appear to have higher rates of short-term vasopressor use, and a greater number of organ failures, contrasted to patients without a sustained SI elevation. An elevated SI may be a useful modality to identify patients with severe sepsis at risk for disease escalation and cardiovascular collapse.[125] Because of these attributes, confirmatory trials of EGDT such as the ProCESS trial used SI once an hour as a resuscitation endpoint.[126]

Capillary refill time (CRT) is a clinical reproducible parameter when measured on the index finger tip or the knee area. After initial resuscitation of septic shock, CRT is a strong predictive factor of 14-day mortality.[127] CRT should not replace the mottling score, but could be used as a complementary tool for several reasons. First, the mottling score is a semi quantitative parameter whereas CRT is a quantitative parameter, leading to more accurate monitoring during resuscitation. Moreover, in the mottling group (0–1) and (2–3), knee CRT improved patient discrimination according to their outcome, with non-survivors presenting a significantly higher knee CRT.

A physical examination that includes evaluation of the skin temperature (cool or warm) of the distal extremities is usually the first step in evaluating the perfusion status. The finding of cool distal extremities upon physical examination, combined with laboratory studies demonstrating low serum bicarbonate and high arterial lactate levels, aid in identifying patients with hypoperfusion.[128] The mottling score is reproducible and easy to evaluate at the bedside. The mottling score using the knee cap areas as well as its variation during resuscitation is a strong predictor of 14-day survival in patients with septic shock.[129,130]

Method	Variable	Advantages	Limitations	Suggested cut-offs for higher mortality
Mottling of the skin	Absence/presence	Could be done by nurses	Lack of specificity	–
	Mottling score	Easy to use and learn reproducible	Not useful in patients with dark skin	Score 4–5 (scoring from 0 to 5)
Capillary refill time (CRT)	Index CRT	Easy to use and learn ± reproducible	Inter rater variability	Critically ill > 5 s Septic shock > 2.4 s
	Knee CRT	Reproducible	Not useful in patients with dark skin	Septic shock > 4.9 s
Temperature gradient	Forearm-to-finger	Validated method	Requires more complex technology	>4 °C
	Central-to-toe	Validated method		>7 °C

Table 1. Methods used to measure peripheral tissue perfusion[131,132]

- **Measure CVP**

Central venous catheter placement has multiple important uses. It provides a safe route for vasopressors and measurement of central venous pressure (CVP) which is an endpoint of fluid administration. It also allows for examination of right heart waveforms (a,c,v with x and y descent) and measurement of metabolic parameters such as venous oxygen saturation ($ScvO_2$), pH, lactate and arterial-venous CO_2 differences.[133-135] CVP guided early volume therapy reduces vasopressor support and the need for corticosteroid therapy.[47,70,73,74,136-138] The use of the central venous catheter is clinically equal to the pulmonary artery catheter in the fluid management of acute lung injury (ALI) and is associated with improved outcomes in severe sepsis and septic shock.[84,139-144]

- **Measure $ScvO_2$**

The use of $ScvO_2$ for the management of shock has been described for over a half century.[145] Many of the salutary effects of $ScvO_2$ are based on its diagnostic ability to detect early imbalances of systemic oxygen delivery (DO_2) to systemic oxygen consumption (VO_2). This is important particularly in the DO_2 dependent phase of sepsis, where vital signs, shock index and even lactate can be normal.[146,147] In animal models of sepsis, a low $ScvO_2$ is uniformly present after the insult, even in sedated and mechanically ventilated subjects at baseline.[148-153] This DO_2 dependent phase (low SvO_2/$ScvO_2$) is frequently seen early in patients both in the ED and ICU setting before interventions.[154-156] In the presence of a low $ScvO_2$ value; therapeutic maneuvers to increase DO_2 or decrease VO_2 are required to prevent tissue hypoxia, inflammation, lactate generation, cardiopulmonary complications and increased mortality.[147,155,157] A low $ScvO_2$ is a trigger for increasing inspired oxygen concentration (arterial hypoxia)[158,159] and fluid therapy for a low CVP.[150,160,161] It is also a trigger for red blood cell transfusion [162] (anemia or decreased hemoglobin content), titrating inotrope therapy (myocardial suppression or strain)[163-166] and the need for mechanical ventilation with sedation (increased oxygen demands).[154,165,167,168] $ScvO_2$ improves with mechanical ventilation and predicts extubation failure.[154,169-171] As a result of its diagnostic and therapeutic utility, the use of $ScvO_2$ is significantly associated with improved outcomes.[16,70,84,156,166,172-180]

Outcome Evidence Since the Last Measure Submission

A significant reduction in sepsis mortality began after the millennium, Figure 1. This coincided with the introduction of seminal studies using components of the current sepsis measure, Figure 2.[191] A recent international expert examination of over 52 studies (166,479 patients between January 1, 1992 and December 25, 2015) revealed this period began with a mortality of 46.5%.[192] This mortality is identical to the control group of the EGDT trial whose findings have been reproduced in multiple trial designs. Of note is the abundance of observational studies, Figure 2 and Table 2.[8] While randomized controlled trials (RCTs) are considered the standard, large prospective observational studies provide an equally reliable scientific alternative to RCTs.[193] In particular, multiple large collaborative quality initiatives in addition to the Surviving Sepsis Campaign have shown significant mortality reduction.[26,194-197] No studies (observational or otherwise) exist in the literature that fail to demonstrate a reduction in mortality when “sepsis bundles” employing all or nothing composite measures have been applied.

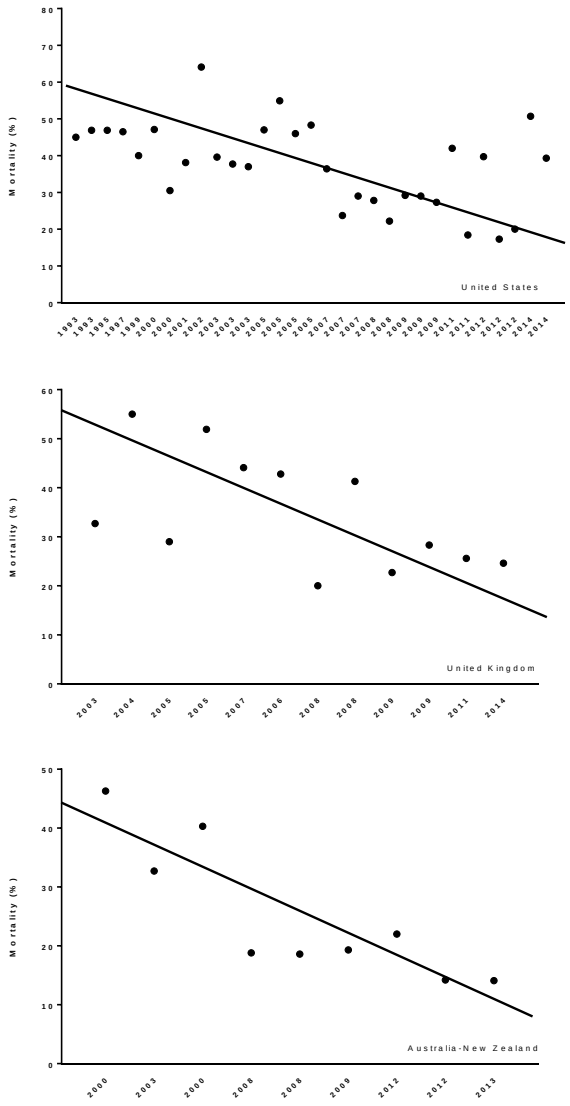


Figure 1. Trending mortality rates of observational studies of severe and septic shock in the United States [126,198-208], United Kingdom [209-216] and Australia and New Zealand [217-222]. This represents a consistent reduction in sepsis mortality over the last two decades.[223,224]

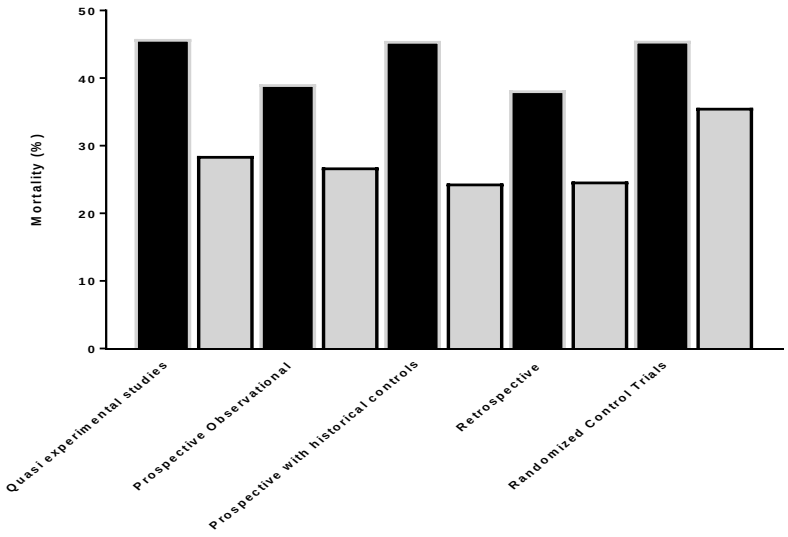


Figure 2. A consistent reduction in mortality irrespective of study design. The black columns are the intervention group and the grey columns are the control or non-intervention groups.[223,224]

Studies	Studies N	Patients N	Control Mortality	Intervention Mortality
Quasi experimental studies [225-228]	4	1120	45.8	28.5
Prospective Observational [47,58,74,180,195-197,210,220,229-267]	44	169764	38.1	26.2
Prospective with historical controls [268-276]	9	2250	45.5	29.6
Retrospective [214,277-286]	12	2616	38.2	24.4
Randomized Control Trials [7,287-300]	12	5756	45.5	35.6

Table 2. N represents the number of studies of severe sepsis and septic shock followed by the total number of patients. The mortality reflects the average of all studies.

Independent Data Analysis

Since the last submission, the NQF-0500 measure was introduced as SEP-1, the first CMS sepsis measure with reporting beginning in October, 2015. The inpatient Quality Reporting (IQR) Program Performance requirements were met by 99% of the reporting hospitals. Using SEP-1 data from the Clinical Data Warehouse (CDW), CMS conducted a time trend analysis to assess the mortality rate across 159, 289 patients who were eligible for the SEP-1 measure over three quarters of reporting. The results of the analysis (Table 3) demonstrate that the mortality rate for patients who received all applicable elements of care for the SEP-1 measure was on average 8.5% lower (21.3% in 2015 Q4, 23.0% in 2016 Q1, 21.4% in 2016 Q2) compared to patients who did not receive all applicable elements of care (29.6%, in 2015 Q4, 31.8% in 2016 Q1, 29.7% in 2016 Q2). This trend was consistent across all time periods of data available for analysis. Additionally, the mortality rates demonstrate the potential impact of the measure in the number of preventable deaths as a result of the SEP-1 measure. The relative mortality reduction observed with the introduction of SEP-1 is similar to that seen with quality initiatives over the last 15 years, Figure 2 and Table 2.

	Severe Sepsis and Septic Shock Mortality Rate		
Description	**2015 Q4**	**2016 Q1**	**2016 Q2**
Did not Meet Guidelines for SEP-1	29.6%	31.8%	29.7%

Met Guidelines for SEP-1	21.3%	23.0%	21.4%
Absolute Reduction Rate	8.3%	8.8%	8.3%
Relative Reduction Rate	28.04%	27.7%	27.9%
	Potential Preventable Deaths		
	2,783	2,864	2,411

Table 3: Sepsis Mortality Analysis and Trends, data source: Clinical Inpatient SEP-1 Measure Cases with Identified Medicare Payment Source

Additional detail on the mortality analysis is included in the testing form.

1a.4.2 What process was used to identify the evidence?

The above elements reflect best practice recommendations from the Surviving Sepsis Campaign.[69] The evidence presented reflects a comprehensive on-line search of sepsis related studies.[223,224,301] All existing meta-analysis of protocolized sepsis care were also reviewed.[63,174,302-321]

1a.4.3. Provide the citation(s) for the evidence.

1. Levy MM, Dellinger RP, Townsend SR, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. *Crit Care Med.* 2010;38(2):367-374.
2. Grzybowski M, Tuttle A, Nowak R, et al. Systemic inflammatory response syndrome criteria and lactic acidosis in the detection of critical illness among patients presenting to the emergency department. *Chest.* 1996;110:145S.
3. Brun-Buisson C, Doyon F, Carlet J, et al. Incidence, risk factors, and outcome of severe sepsis and septic shock in adults. A multicenter prospective study in intensive care units. French ICU Group for Severe Sepsis. *Jama.* 1995;274(12):968-974.
4. Rivers EP, Kruse JA, Jacobsen G, et al. The influence of early hemodynamic optimization on biomarker patterns of severe sepsis and septic shock. *Crit Care Med.* 2007;35(9):2016-2024.
5. Keegan J, Wira CR. Early Identification and Management of Patients with Severe Sepsis and Septic Shock in the Emergency Department. *Emergency Medicine Clinics of North America.* 2014;32(4):759-+.
6. Carr GE, Yuen TC, McConville JF, et al. Early Cardiac Arrest in Patients Hospitalized With Pneumonia: A Report From the American Heart Association's Get With the Guidelines-Resuscitation Program. *Chest.* 2012;141(6):1528-1536.
7. Lin SM, Huang CD, Lin HC, Liu CY, Wang CH, Kuo HP. A modified goal-directed protocol improves clinical outcomes in intensive care unit patients with septic shock: a randomized controlled trial. *Shock.* 2006;26(6):551-557.
8. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med.* 2001;345(19):1368-1377.
9. Levy MM, Macias WL, Vincent JL, et al. Early changes in organ function predict eventual survival in severe sepsis. *Crit Care Med.* 2005;33(10):2194-2201.
10. Estenssoro E, Gonzalez F, Laffaire E, et al. Shock on admission day is the best predictor of prolonged mechanical ventilation in the ICU. *Chest.* 2005;127(2):598-603.
11. Goyal N, Taylor AR, Rivers EP. Relationship between Central and Peripheral Venous Oxygen Saturation and Lactate Levels: A Prospective Study. *Journal of Emergency Medicine.* 2016.
12. Reminiac F, Saint-Etienne C, Runge I, et al. Are central venous lactate and arterial lactate interchangeable? A human retrospective study. *Anesth Analg.* 2012;115(3):605-610.
13. Puskarich MA, Trzeciak S, Shapiro NI, et al. Whole blood lactate kinetics in patients undergoing quantitative resuscitation for severe sepsis and septic shock. *Chest.* 2013;143(6):1548-1553.
14. Howell MD, Donnino M, Clardy P, Talmor D, Shapiro NI. Occult hypoperfusion and mortality in patients with suspected infection. *Intensive Care Med.* 2007;33(11):1892-1899.
15. Singer AJ, Taylor M, Domingo A, et al. Diagnostic characteristics of a clinical screening tool in combination with measuring bedside lactate level in emergency department patients with suspected sepsis. *Acad Emerg Med.* 2014;21(8):853-857.

16. Shin TG, Jo IJ, Hwang SY, et al. Comprehensive Interpretation of Central Venous Oxygen Saturation and Blood Lactate Levels During Resuscitation of Patients With Severe Sepsis and Septic Shock in the Emergency Department. *Shock.* 2016;45(1):4-9.
17. Varis E, Pettila V, Poukkanen M, et al. Evolution of Blood Lactate and 90-Day Mortality in Septic Shock. A Post Hoc Analysis of the Finnaki Study. *Shock.* 2016.
18. Broder G, Weil MH. Excess lactate: an index of irreversibility of shock in human patients. *Science.* 1964;143(3613):1457-1459.
19. Cady LD, Jr., Weil MH, Afifi AA, Michaels SF, Liu VY, Shubin H. Quantitation of severity of critical illness with special reference to blood lactate. *Crit Care Med.* 1973;1(2):75-80.
20. Aduen J, Bernstein WK, Khastgir T, et al. The use and clinical importance of a substrate-specific electrode for rapid determination of blood lactate concentrations. *JAMA.* 1994;272(21):1678-1685.
21. Mikkelsen ME, Miltiades AN, Gaieski DF, et al. Serum lactate is associated with mortality in severe sepsis independent of organ failure and shock. *Crit Care Med.* 2009;37(5):1670-1677.
22. Trzeciak S, Dellinger RP, Chansky ME, et al. Serum lactate as a predictor of mortality in patients with infection. *Intensive Care Med.* 2007;33(6):970-977.
23. Shapiro NI, Howell MD, Talmor D, et al. Serum lactate as a predictor of mortality in emergency department patients with infection. *Ann Emerg Med.* 2005;45(5):524-528.
24. Pearse RM. Extending the role of lactate measurement into the prehospital environment. *Crit Care.* 2009;13(1):115.
25. Casserly B, Phillips GS, Schorr C, et al. Lactate measurements in sepsis-induced tissue hypoperfusion: results from the surviving sepsis campaign database. *Crit Care Med.* 2015;43(3):567-573.
26. Liu VX, Morehouse JW, Marelich GP, et al. Multicenter Implementation of a Treatment Bundle for Patients with Sepsis and Intermediate Lactate Values. *Am J Respir Crit Care Med.* 2016;193(11):1264-1270.
27. Puskarich MA, Illich BM, Jones AE. Prognosis of emergency department patients with suspected infection and intermediate lactate levels: A systematic review. *Journal of Critical Care.* 2014;29(3):334-339.
28. Song YH, Shin TG, Kang MJ, et al. Predicting factors associated with clinical deterioration of sepsis patients with intermediate levels of serum lactate. *Shock.* 2012;38(3):249-254.
29. Leisman DE, Zemmel D'Amore JA, Gribben JL, et al. Early sepsis bundle compliance for non-hypotensive patients with intermediate versus severe hyperlactemia. *The American Journal of Emergency Medicine.* 2017.
30. Singer AJ, Taylor M, LeBlanc D, Williams J, Thode HC, Jr. ED bedside point-of-care lactate in patients with suspected sepsis is associated with reduced time to iv fluids and mortality. *Am J Emerg Med.* 2014;32(9):1120-1124.
31. Arnold R, Zhang Z, Patel S, Smola S, Isserman J, Jackson E. Delayed assessment of serum lactate in sepsis is associated with an increased mortality rate. *Critical Care.* 2014;18(Suppl 1):P174.
32. Hwang SY, Shin TG, Jo IJ, et al. Association between Hemodynamic Presentation and Outcome in Sepsis Patients. *Shock.* 2014;42(3):205-210.
33. Whittaker SA, Fuchs BD, Gaieski DF, et al. Epidemiology and outcomes in patients with severe sepsis admitted to the hospital wards. *J Crit Care.* 2015;30(1):78-84.
34. Jones AE, Shapiro NI, Trzeciak S, Arnold RC, Claremont HA, Kline JA. Lactate clearance vs central venous oxygen saturation as goals of early sepsis therapy: a randomized clinical trial. *JAMA.* 2010;303(8):739-746.
35. Pro CI, Yealy DM, Kellum JA, et al. A randomized trial of protocol-based care for early septic shock. *N Engl J Med.* 2014;370(18):1683-1693.
36. Riedel S, Bourbeau P, Swartz B, et al. Timing of specimen collection for blood cultures from febrile patients with bacteremia. *J Clin Microbiol.* 2008;46(4):1381-1385.
37. Baron EJ, Miller JM, Weinstein MP, et al. A guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2013 recommendations by the Infectious Diseases Society of America (IDSA) and the American Society for Microbiology (ASM)(a). *Clin Infect Dis.* 2013;57(4):e22-e121.
38. O'Grady NP, Alexander M, Burns LA, et al. Guidelines for the Prevention of Intravascular Catheter-related Infections. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America.* 2011;52(9):e162-e193.

39. Mermel LA, Allon M, Bouza E, et al. Clinical practice guidelines for the diagnosis and management of intravascular catheter-related infection: 2009 Update by the Infectious Diseases Society of America. *Clinical infectious diseases.* 2009;49(1):1-45.
40. Boyce JM, Nadeau J, Dumigan D, et al. Obtaining blood cultures by venipuncture versus from central lines: Impact on blood culture contamination rates and potential effect on central line-associated bloodstream infection reporting. *Infection Control and Hospital Epidemiology.* 2013;34(10):1042-1047.
41. Beekmann SE, Diekema DJ, Charles Huskins W, et al. Diagnosing and reporting of central line-associated bloodstream infections. *Infection Control and Hospital Epidemiology.* 2012;33(9):875-882.
42. Zadroga R, Williams DN, Gottschall R, et al. Comparison of 2 blood culture media shows significant differences in bacterial recovery for patients on antimicrobial therapy. *Clinical Infectious Diseases.* 2013;56(6):790-797.
43. Kanegaye JT, Soliemanzadeh P, Bradley JS. Lumbar puncture in pediatric bacterial meningitis: Defining the time interval for recovery of cerebrospinal fluid pathogens after parenteral antibiotic pretreatment. *Pediatrics.* 2001;108(5):1169-1174.
44. Kumar A, Roberts D, Wood KE, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med.* 2006;34(6):1589-1596.
45. Long B, Koyfman A. Best Clinical Practice: Blood Culture Utility in the Emergency Department. *J Emerg Med.* 2016;51(5):529-539.
46. Filbin MR, Arias SA, Camargo CA, Jr., Barche A, Pallin DJ. Sepsis visits and antibiotic utilization in u.s. Emergency departments*. *Crit Care Med.* 2014;42(3):528-535.
47. Cardoso T, Carneiro AH, Ribeiro O, Teixeira-Pinto A, Costa-Pereira A. Reducing mortality in severe sepsis with the implementation of a core 6-hour bundle: results from the Portuguese community-acquired sepsis study (SACiUCI study). *Crit Care.* 2010;14(3):R83.
48. Kang CI, Song JH, Kim SH, et al. Risk factors and pathogenic significance of bacteremic pneumonia in adult patients with community-acquired pneumococcal pneumonia. *J Infect.* 2012.
49. Kumar A, Ellis P, Arabi Y, et al. Initiation of Inappropriate Antimicrobial Therapy Results in a 5-Fold Reduction of Survival in Human Septic Shock. *Chest.* 2009;136(5):1237-1248.
50. Lindvig KP, Nielsen SL, Henriksen DP, et al. Mortality and prognostic factors of patients who have blood cultures performed in the emergency department: a cohort study. *European Journal of Emergency Medicine.* 2016;23(3):166-172.
51. Savage RD, Fowler RA, Rishu AH, et al. The Effect of Inadequate Initial Empiric Antimicrobial Treatment on Mortality in Critically Ill Patients with Bloodstream Infections: A Multi-Centre Retrospective Cohort Study. *PLoS One.* 2016;11(5):e0154944.
52. Garnacho-Montero J, Garcia-Garmendia JL, Barrero-Almodovar A, Jimenez-Jimenez FJ, Perez-Paredes C, Ortiz-Leyba C. Impact of adequate empirical antibiotic therapy on the outcome of patients admitted to the intensive care unit with sepsis. *Crit Care Med.* 2003;31(12):2742-2751.
53. Cardoso T, Carneiro AH, Ribeiro O, Teixeira-Pinto A, Costa-Pereira A. Reducing mortality in severe sepsis with the implementation of a core 6-h bundle: results from the Portuguese community-acquired sepsis study (SACiUCI study). *Crit Care.* 2010;14.
54. de Sousa AG, Junior CJF, Santos GPD, et al. The impact of each action in the Surviving Sepsis Campaign measures on hospital mortality of patients with severe sepsis/septic shock. *Einstein.* 2008;6(3):323-327.
55. Garnacho-Montero J, Gutierrez-Pizarraya A, Escoresca-Ortega A, et al. De-escalation of empirical therapy is associated with lower mortality in patients with severe sepsis and septic shock. *Intensive Care Med.* 2014;40(1):32-40.
56. Ferrer R, Martin-Loeches I, Phillips G, et al. Empiric Antibiotic Treatment Reduces Mortality in Severe Sepsis and Septic Shock From the First Hour: Results From a Guideline-Based Performance Improvement Program. *Crit Care Med.* 2014;42(8):1749-1755.
57. Natanson C, Danner RL, Reilly JM, et al. Antibiotics versus cardiovascular support in a canine model of human septic shock. *Am J Physiol.* 1990;259(5 Pt 2):H1440-1447.
58. Ferrer R, Artigas A, Levy MM, et al. Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain. *JAMA.* 2008;299(19):2294-2303.

59. Siddiqui S, Razzak J. Early versus late pre-intensive care unit admission broad spectrum antibiotics for severe sepsis in adults. *Cochrane Database Syst Rev.* 2010;10:CD007081.
60. Gaieski DF, Mikkelsen ME, Band RA, et al. Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. *Crit Care Med.* 2010;38(4):1045-1053.
61. Puskarich M, Trzeciak S, Shapiro N, et al. Association between timing of antibiotic administration and mortality from septic shock in patients treated with a quantitative resuscitation protocol. *Crit Care Med.* 2011;39:2066 - 2071.
62. Hutchison RW, Govathoti DA, Fehlis K, et al. Improving severe sepsis outcomes: cost and time to first antibiotic dose. *Dimens Crit Care Nurs.* 2011;30(5):277-282.
63. Barochia AV, Cui X, Vitberg D, et al. Bundled care for septic shock: an analysis of clinical trials. *Crit Care Med.* 2010;38(2):668-678.
64. Judd W, Stephens D, Kennedy C. Clinical and Economic Impact of a Quality Improvement Initiative to Enhance Early Recognition and Treatment of Sepsis. *Ann Pharmacother.* 2014;48(10):1269-1275.
65. Houck PM, Bratzler DW, Nsa W, Ma A, Bartlett JG. Timing of antibiotic administration and outcomes for medicare patients hospitalized with community-acquired pneumonia. *Arch Intern Med.* 2004;164(6):637-644.
66. Garnacho-Montero J, Ortiz-Leyba C, Herrera-Melero I, et al. Mortality and morbidity attributable to inadequate empirical antimicrobial therapy in patients admitted to the ICU with sepsis: a matched cohort study. *J Antimicrob Chemother.* 2008;61(2):436-441.
67. Capp R, Chang Y, Brown DF. Effective antibiotic treatment prescribed by emergency physicians in patients admitted to the intensive care unit with severe sepsis or septic shock: where is the gap? *J Emerg Med.* 2011;41(6):573-580.
68. Dellinger RP, Levy MM, Rhodes A, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med.* 2013;41.
69. Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Crit Care Med.* 2017.
70. Lee SJ, Ramar K, Park JG, Gajic O, Li G, Kashyap R. Increased fluid administration in the first three hours of sepsis resuscitation is associated with reduced mortality: A retrospective cohort study. *Chest.* 2014;146(4):908-915.
71. Jones AE, Yiannibas V, Johnson C, Kline JA. Emergency department hypotension predicts sudden unexpected in-hospital mortality: a prospective cohort study. *Chest.* 2006;130(4):941-946.
72. Leisman D, Wie B, Doerfler M, et al. Association of Fluid Resuscitation Initiation Within 30 Minutes of Severe Sepsis and Septic Shock Recognition With Reduced Mortality and Length of Stay. *Ann Emerg Med.* 2016;68(3):298-311.
73. Practice parameters for hemodynamic support of sepsis in adult patients in sepsis. Task Force of the American College of Critical Care Medicine, Society of Critical Care Medicine. *Crit Care Med.* 1999;27(3):639-660.
74. Micek ST, Roubinian N, Heuring T, et al. Before-after study of a standardized hospital order set for the management of septic shock. *Crit Care Med.* 2006;34(11):2707-2713.
75. Dorresteijn MJ, van Eijk LT, Netea MG, Smits P, van der Hoeven JG, Pickkers P. Iso-osmolar prehydration shifts the cytokine response towards a more anti-inflammatory balance in human endotoxemia. *J Endotoxin Res.* 2005;11(5):287-293.
76. Jahoor A, Delmas T, Giri B, et al. FLuid resuscitation of at least 1 liter in septic patients decreases the need for renal replacement therapy without increasing the risk of acute congestive heart failure or acute respiratory failure. *Chest.* 2016;150(4_S):349A-349A.
77. Tam WS, Fried J, Grotts J, Paras T. FLuid resuscitation and clinical outcomes in patients with known heart failure who develop severe sepsis or septic shock. *Chest.* 2016;150(4_S):350A-350A.
78. Murphy CV, Schramm GE, Doherty JA, et al. The Importance of Fluid Management in Acute Lung Injury Secondary to Septic Shock. *Chest.* 2009.
79. Morehouse J, Marelich G, Soule J, et al. Mortality Among Sepsis Patients With Intermediate Lactate Values Following Multicenter Implementation Of A Treatment Bundle. *Am J Respir Crit Care Med.* 2015;191:A4004.
80. Suarez D, Ferrer R, Artigas A, et al. Cost-effectiveness of the Surviving Sepsis Campaign protocol for severe sepsis: a prospective nation-wide study in Spain. *Intensive Care Med.* 2011;37(3):444-452.
81. Dunser MW, Takala J, Ulmer H, et al. Arterial blood pressure during early sepsis and outcome. *Intensive Care Med.* 2009;35(7):1225-1233.

82. Beck V, Chateau D, Bryson GL, et al. Timing of vasopressor initiation and mortality in septic shock: a cohort study. *Critical Care.* 2014;18(3):R97.
83. Thooft A, Favory R, Salgado DR, et al. Effects of changes in arterial pressure on organ perfusion during septic shock. *Crit Care.* 2011;15(5):R222.
84. Varpula M, Tallgren M, Saukkonen K, Voipio-Pulkki LM, Pettila V. Hemodynamic variables related to outcome in septic shock. *Intensive Care Med.* 2005;31:1066-1071.
85. Asfar P, Meziani F, Hamel J-F, et al. High versus Low Blood-Pressure Target in Patients with Septic Shock. *New England Journal of Medicine.* 2014;370(17):1583-1593.
86. Leone M, Asfar P, Radermacher P, Vincent J-L, Martin C. Optimizing mean arterial pressure in septic shock: a critical reappraisal of the literature. *Critical Care.* 2015;19(1):101.
87. Beloncle F, Radermacher P, Guerin C, Asfar P. Mean arterial pressure target in patients with septic shock. *Minerva Anestesiol.* 2016.
88. Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Medicine.* 2017:1-74.
89. O'Neill R, Morales J, Jule M. Early goal-directed therapy (EGDT) for severe sepsis/septic shock: which components of treatment are more difficult to implement in a community-based emergency department? *J Emerg Med.* 2012;42(5):503-510.
90. Labovitz AJ, Noble VE, Bierig M, et al. Focused cardiac ultrasound in the emergent setting: a consensus statement of the American Society of Echocardiography and American College of Emergency Physicians. *J Am Soc Echocardiogr.* 2010;23(12):1225-1230.
91. Arntfield RT, Millington SJ. Point of Care Cardiac Ultrasound Applications in the Emergency Department and Intensive Care Unit - A Review. *Current Cardiology Reviews.* 2012;8(2):98-108.
92. Perera P, Mailhot T, Riley D, Mandavia D. The RUSH exam: Rapid Ultrasound in SHock in the evaluation of the critically lll. *Emerg Med Clin North Am.* 2010;28(1):29-56, vii.
93. Chew HC, Devanand A, Phua GC, Loo CM. Oesophageal Doppler ultrasound in the assessment of haemodynamic status of patients admitted to the medical intensive care unit with septic shock. *Ann Acad Med Singapore.* 2009;38(8):699-703.
94. Beaulieu Y. Bedside echocardiography in the assessment of the critically ill. *Critical Care Medicine.* 2007;35(5 Suppl):S235-249.
95. Guérin L, Vieillard-Baron A. The Use of Ultrasound in Caring for Patients with Sepsis. *Clinics in chest medicine.* 2016;37(2):299-307.
96. Coen D, Vaccaro A, Cazzaniga M, Cortellaro F, Monti G, Tombini V. Toward a noninvasive approach to early goal-directed therapy. *Chest.* 2011;139(3):726-727.
97. Shokoohi H, Boniface KS, Pourmand A, et al. Bedside Ultrasound Reduces Diagnostic Uncertainty and Guides Resuscitation in Patients With Undifferentiated Hypotension. *Crit Care Med.* 2015;43(12):2562-2569.
98. Jones AE, Tayal VS, Sullivan DM, Kline JA. Randomized, controlled trial of immediate versus delayed goal-directed ultrasound to identify the cause of nontraumatic hypotension in emergency department patients. *Crit Care Med.* 2004;32(8):1703-1708.
99. Smith MD, MacPhail B, Harrison MR, Lenhoff SJ, DeMaria AN. Value and limitations of transesophageal echocardiography in determination of left ventricular volumes and ejection fraction. *J Am Coll Cardiol.* 1992;19(6):1213-1222.
100. Gunn SR, Pinsky MR. Implications of arterial pressure variation in patients in the intensive care unit. *Current opinion in critical care.* 2001;7(3):212-217.
101. Keller AS, Melamed R, Malinchoc M, John R, Tierney DM, Gajic O. Diagnostic accuracy of a simple ultrasound measurement to estimate central venous pressure in spontaneously breathing, critically ill patients. *J Hosp Med.* 2009;4(6):350-355.
102. Worapratya P, Anupat S, Suwannanon R, Wuthisuthimethawee P. Correlation of caval index, inferior vena cava diameter, and central venous pressure in shock patients in the emergency room. *Open Access Emerg Med.* 2014;6:57-62.
103. Prekker ME, Scott NL, Hart D, Sprenkle MD, Leatherman JW. Point-of-care ultrasound to estimate central venous pressure: a comparison of three techniques. *Crit Care Med.* 2013;41(3):833-841.

104. Bentzer P, Griesdale DE, Boyd J, MacLean K, Sirounis D, Ayas NT. Will This Hemodynamically Unstable Patient Respond to a Bolus of Intravenous Fluids? *JAMA.* 2016;316(12):1298-1309.
105. Lanspa MJ, Pittman JE, Hirshberg EL, et al. Association of left ventricular longitudinal strain with central venous oxygen saturation and serum lactate in patients with early severe sepsis and septic shock. *Crit Care.* 2015;19(1):1-9.
106. Via G, Storti E, Spreafico A, Melniker L, Neri L. Point of care ultrasound for sepsis management in resource-limited settings: time for a new paradigm for global health care. *Intensive Care Med.* 2012;38(8):1405-1407; author reply 1408-1409.
107. Mehta MH, Webster W, Hynes DM. Intra-abdominal sepsis--ultrasound evaluation. *J Can Assoc Radiol.* 1978;29(2):122-123.
108. Rozycki GS. Surgeon-performed ultrasound: its use in clinical practice. *Annals of Surgery.* 1998;228(1):16-28.
109. Garg M, Sen J, Goyal S, Chaudhry D. Comparative evaluation of central venous pressure and sonographic inferior vena cava variability in assessing fluid responsiveness in septic shock. *Indian Journal of Critical Care Medicine.* 2016;20(12):708-713.
110. Shiver S, Blaivas M, Lyon M. A prospective comparison of ultrasound-guided and blindly placed radial arterial catheters. *Acad Emerg Med.* 2006;13(12):1275-1279.
111. Milling T, Holden C, Melniker L, Briggs WM, Birkhahn R, Gaeta T. Randomized controlled trial of single-operator vs. two-operator ultrasound guidance for internal jugular central venous cannulation. *Acad Emerg Med.* 2006;13(3):245-247.
112. Cortellaro F, Colombo S, Coen D, Duca PG. Lung ultrasound is an accurate diagnostic tool for the diagnosis of pneumonia in the emergency department. *Emerg Med J.* 2012;29(1):19-23.
113. Kehrl T. Point-of-care ultrasound diagnosis of necrotizing fasciitis missed by computed tomography and magnetic resonance imaging. *J Emerg Med.* 2014;47(2):172-175.
114. Mongodi S, Via G, Girard M, et al. Lung Ultrasound for Early Diagnosis of Ventilator-Associated Pneumonia. *Chest.* 2016;149(4):969-980.
115. Cortellaro F, Ferrari L, Molteni F, et al. Accuracy of point of care ultrasound to identify the source of infection in septic patients: a prospective study. *Internal and emergency medicine.* 2016:1-8.
116. Akhtar S, Theodoro D, Gaspari R, et al. Resident training in emergency ultrasound: consensus recommendations from the 2008 Council of Emergency Medicine Residency Directors Conference. *Acad Emerg Med.* 2009;16 Suppl 2:S32-36.
117. Hyzy RC. Bedside Ultrasonography for Diagnosis of Septic Shock. *JAMA.* 2016;315(1):89.
118. Michard F, Chemla D, Richard C, et al. Clinical use of respiratory changes in arterial pulse pressure to monitor the hemodynamic effects of PEEP. *Am J Respir Crit Care Med.* 1999;159(3):935-939.
119. Michard F, Lopes M, Auler J-O. Pulse pressure variation: beyond the fluid management of patients with shock. *Critical Care.* 2007;11(3):131.
120. Rady MY. The role of central venous oximetry, lactic acid concentration and shock index in the evaluation of clinical shock: a review. *Resuscitation.* 1992;24(1):55-60.
121. Berger T, Green J, Horeczko T, et al. Shock index and early recognition of sepsis in the emergency department: pilot study. *West J Emerg Med.* 2013;14(2):168-174.
122. Rady MY, Nightingale P, Little RA, Edwards JD. Shock index: a re-evaluation in acute circulatory failure. *Resuscitation.* 1992;23(3):227-234.
123. Rady MY, Rivers EP, Martin GB, Smithline H, Appelton T, Nowak RM. Continuous central venous oximetry and shock index in the emergency department: use in the evaluation of clinical shock. *Am J Emerg Med.* 1992;10(6):538-541.
124. Rady MY, Smithline HA, Blake H, Nowak R, Rivers E. A comparison of the shock index and conventional vital signs to identify acute, critical illness in the emergency department. *Ann Emerg Med.* 1994;24(4):685-690.
125. Wira CR, Francis MW, Bhat S, Ehrman R, Conner D, Siegel M. The shock index as a predictor of vasopressor use in emergency department patients with severe sepsis. *West J Emerg Med.* 2014;15(1):60-66.
126. Investigators P, Yealy DM, Kellum JA, et al. A randomized trial of protocol-based care for early septic shock. *N Engl J Med.* 2014;370(18):1683-1693.

127. Ait-Oufella H, Bige N, Boelle PY, et al. Capillary refill time exploration during septic shock. *Intensive Care Med.* 2014;40(7):958-964.
128. Kaplan LJ, McPartland K, Santora TA, Trooskin SZ. Start with a subjective assessment of skin temperature to identify hypoperfusion in intensive care unit patients. *J Trauma.* 2001;50(4):620-627; discussion 627-628.
129. Ait-Oufella H, Lemoinne S, Boelle PY, et al. Mottling score predicts survival in septic shock. *Intensive Care Med.* 2011;37(5):801-807.
130. Galbois A, Bige N, Pichereau C, et al. Exploration of skin perfusion in cirrhotic patients with septic shock. *Journal of hepatology.* 2015;62(3).
131. Ait-Oufella H, Bakker J. Understanding clinical signs of poor tissue perfusion during septic shock. *Intensive Care Med.* 2016;42(12):2070-2072.
132. van Genderen ME, van Bommel J, Lima A. Monitoring peripheral perfusion in critically ill patients at the bedside. *Current opinion in critical care.* 2012;18(3):273-279.
133. Landis EM, Brown E, et al. Central venous pressure in relation to cardiac competence, blood volume and exercise. *J Clin Invest.* 1946;25:237-255.
134. Weil M, Shubin H, Rosoff L. Fluid repletion in circulatory shock: Central venous pressure and other practical guides. *JAMA.* 1965;192(8):668-674.
135. Cuscheuri J, Katilius M, Hays G, Rivers E. Arterial-Venous Carbon Dioxide Gradients as an Indicator of Cardiac Index: A Comparison Between the Mixed and Central Venous Circulation. *Crit Care Med.* 1998;26(1):A56.
136. Kampmeier TG, Rehberg S, Westphal M, Lange M. Vasopressin in sepsis and septic shock. *Minerva Anestesiol.* 2010;76(10):844-850.
137. Cohen R. Use of corticosteroids in septic shock. *Minerva Anestesiol.* 2011;77(2):190-195.
138. Lee YI, Smith RL, Gartshteyn Y, Kwon S, Caraher EJ, Nolan A. Predictors of Acute Hemodynamic Decompensation in Early Sepsis: An Observational Study. *Journal of Clinical Medicine Research.* 2016;8(8):575-581.
139. Lichtwarck-Aschoff M, Zeravik J, Pfeiffer UJ. Intrathoracic blood volume accurately reflects circulatory volume status in critically ill patients with mechanical ventilation. *Intensive care medicine.* 1992;18(3):142-147.
140. Kashiouris M, Kashyap R, Jaffer-Sathick I, Gajic O, Cartin-Ceba R. Association between delays in central venous catheter utilization and organ dysfunction in patients with severe sepsis and septic shock. *Crit Care Med.* 2011;39((12) Supplement):142.
141. Ferrer R, Artigas A, Suarez D, et al. Effectiveness of treatments for severe sepsis: a prospective multicenter observational study. *Am J Respir Crit Care Med.* 2009;180:861 - 866.
142. Hayatdavoudi S, Crowell R, Rincon T, Winchell J, Yee A, Ikeda D. 505: Does Aggressive Fluid Resucitation Using Cvp Monitoring Decrease Mortality Below the Apache IV Predicted Outcomes? *Critical Care Medicine.* 2012;40(12):1-328 310.1097/1001.ccm.0000424723.0000467161.de.
143. Walkey AJ, Wiener RS, Lindenauer PK. Utilization patterns and outcomes associated with central venous catheter in septic shock: a population-based study. *Crit Care Med.* 2013;41(6):1450-1457.
144. Theodoro D, Owens PL, Olsen MA, Fraser V. Rates and timing of central venous cannulation among patients with sepsis and respiratory arrest admitted by the emergency department*. *Crit Care Med.* 2014;42(3):554-564.
145. Wilson JN. Rational approach to management of clinical shock: Utilizing light-reflection oximetry and central venous pressure monitoring. *Archives of Surgery.* 1965;91(1):92-20.
146. Grissom CK, Morris AH, Lanken PN, et al. Association of physical examination with pulmonary artery catheter parameters in acute lung injury. *Crit Care Med.* 2009;37(10):2720-2726.
147. Krafft P, Steltzer H, Hiesmayr M, Klimscha W, Hammerle AF. Mixed venous oxygen saturation in critically ill septic shock patients. The role of defined events. *Chest.* 1993;103(3):900-906.
148. Rosario AL, Park M, Brunialti MK, et al. SvO(2)-guided resuscitation for experimental septic shock: effects of fluid infusion and dobutamine on hemodynamics, inflammatory response, and cardiovascular oxidative stress. *Shock.* 2011;36(6):604-612.
149. Regueira T, Djafarzadeh S, Brandt S, et al. Oxygen transport and mitochondrial function in porcine septic shock, cardiogenic shock, and hypoxaemia. *Acta Anaesthesiol Scand.* 2012;56(7):846-859.
150. Carroll GC, Snyder JV. Hyperdynamic severe intravascular sepsis depends on fluid administration in cynomolgus monkey. *Am J Physiol.* 1982;243(1):R131-141.

151. Correa TD, Vuda M, Blaser AR, et al. Effect of treatment delay on disease severity and need for resuscitation in porcine fecal peritonitis. *Crit Care Med.* 2012;40(10):2841-2849.
152. Osthaus WA, Huber D, Beck C, et al. Correlation of oxygen delivery with central venous oxygen saturation, mean arterial pressure and heart rate in piglets. *Paediatr Anaesth.* 2006;16(9):944-947.
153. Young BC, Prittie JE, Fox P, Barton LJ. Decreased central venous oxygen saturation despite normalization of heart rate and blood pressure post shock resuscitation in sick dogs. *Journal of Veterinary Emergency and Critical Care.* 2014;24(2):154-161.
154. Hernandez G, Pena H, Cornejo R, et al. Impact of emergency intubation on central venous oxygen saturation in critically ill patients: a multicenter observational study. *Crit Care.* 2009;13(3):R63.
155. Astiz ME, Rackow EC, Kaufman B, Falk JL, Weil MH. Relationship of oxygen delivery and mixed venous oxygenation to lactic acidosis in patients with sepsis and acute myocardial infarction. *Crit Care Med.* 1988;16(7):655-658.
156. Boulain T, Garot D, Vignon P, et al. Prevalence of low central venous oxygen saturation in the first hours of intensive care unit admission and associated mortality in septic shock patients: a prospective multicentre study. *Crit Care.* 2014;18(6):609.
157. Kasnitz P, Druger GL, Yorra F, Simmons DH. Mixed venous oxygen tension and hyperlactatemia. Survival in severe cardiopulmonary disease. *JAMA.* 1976;236(6):570-574.
158. Legrand M, Vallee F, Mateo J, Payen D. Influence of arterial dissolved oxygen level on venous oxygen saturation: don't forget the PaO2! *Shock.* 2014;41(6):510-513.
159. Ho KM, Harding R, Chamberlain J. The impact of arterial oxygen tension on venous oxygen saturation in circulatory failure. *Shock.* 2008;29(1):3-6.
160. Xu B, Yang X, Wang C, et al. Changes of central venous oxygen saturation define fluid responsiveness in patients with septic shock: A prospective observational study. *J Crit Care.* 2016;38:13-19.
161. Xu B, Yang X, Wang C, et al. Changes of central venous oxygen saturation define fluid responsiveness in patients with septic shock: A prospective observational study. *Journal of Critical Care.*38:13-19.
162. Vallet B, Robin E, Lebuffe G. Venous oxygen saturation as a physiologic transfusion trigger. *Crit Care.* 2010;14(2):213.
163. Hirshberg E, Grissom C, Wilson E, Knox D, Brown S, Lanspa M. Decreased Left Ventricular Strain Is Associated With Central Venous Oxygenation Saturation In Patients With Severe Sepsis And Septic Shock. *Am J Respir Crit Care Med.* 2014;189:A5499.
164. Mallat J, Benzidi Y, Salleron J, et al. Time course of central venous-to-arterial carbon dioxide tension difference in septic shock patients receiving incremental doses of dobutamine. *Intensive Care Med.* 2014;40(3):404-411.
165. Perner A, Haase N, Wiis J, White JO, Delaney A. Central venous oxygen saturation for the diagnosis of low cardiac output in septic shock patients. *Acta Anaesthesiol Scand.* 2010;54(1):98-102.
166. Ouellette DR, Shah SZ. Comparison of outcomes from sepsis between patients with and without pre-existing left ventricular dysfunction: a case-control analysis. *Crit Care.* 2014;18(2):R79.
167. Maddirala S, Khan A. Optimizing hemodynamic support in septic shock using central and mixed venous oxygen saturation. *Crit Care Clin.* 2010;26(2):323-333, table of contents.
168. Walley KR. Use of central venous oxygen saturation to guide therapy. *Am J Respir Crit Care Med.* 2011;184(5):514-520.
169. Stahl W, Radermacher P, Georgieff M, Bracht H. Central venous oxygen saturation and emergency intubation--another piece in the puzzle? *Crit Care.* 2009;13(4):172.
170. Wratney AT. Central venous saturation as a predictor of extubation failure. *Crit Care Med.* 2010;38(2):708-709.
171. Teixeira C, da Silva NB, Savi A, et al. Central venous saturation is a predictor of reintubation in difficult-to-wean patients. *Crit Care Med.* 2010;38(2):491-496.
172. Bracht H, Hänggi M, Jeker B, et al. Incidence of low central venous oxygen saturation during unplanned admissions in a multidisciplinary intensive care unit: an observational study. *Crit Care.* 2007;11(1):R2.
173. Castellanos-Ortega A, Suberviola B, Garcia-Astudillo LA, et al. Impact of the surviving sepsis campaign protocols on hospital length of stay and mortality in septic shock patients: Results of a 3-year follow-up quasi-experimental study. *Crit Care Med.* 2010;38(4):1036-1043.
174. Chamberlain D, Willis E, Bersten A. The severe sepsis bundles as processes of care: a meta-analysis. *Aust Crit Care.* 2011;24(4):229 - 243.

175. Jeong SJ, Song YG, Kim CO, et al. Measurement of Plasma sTREM-1 in Patients with Severe Sepsis Receiving Early Goal-Directed Therapy and Evaluation of Its Usefulness. *Shock.* 2012.
176. Ryoo SM, Kim WY, Sohn CH, et al. Prognostic Value of Timing of Antibiotic Administration in Patients With Septic Shock Treated With Early Quantitative Resuscitation in Emergency Departments. *Am J Med Sci.* 2015;349(4):328-333.
177. Wang H, Cui N, Su L, et al. Prognostic value of extravascular lung water and its potential role in guiding fluid therapy in septic shock after initial resuscitation. *Journal of Critical Care.* 2016;33:106-113.
178. Du W, Liu DW, Wang XT, et al. Combining central venous-to-arterial partial pressure of carbon dioxide difference and central venous oxygen saturation to guide resuscitation in septic shock. *J Crit Care.* 2013;28(6):1110 e1111-1115.
179. Pope JV, Jones AE, Gaieski DF, Arnold RC, Trzeciak S, Shapiro NI. Multicenter study of central venous oxygen saturation (ScvO(2)) as a predictor of mortality in patients with sepsis. *Ann Emerg Med.* 2010;55(1):40-46 e41.
180. Herrán-Monge R, Muriel-Bombín A, García-García MM, et al. Mortality reduction and long-term compliance with Surviving Sepsis Campaign: A nationwide multicenter study. *Shock.* 2016;Publish Ahead of Print.
181. De Backer D, Creteur J, Dubois MJ, et al. The effects of dobutamine on microcirculatory alterations in patients with septic shock are independent of its systemic effects. *Crit Care Med.* 2006;34(2):403-408.
182. Falk JL, Rackow EC, Leavy J, Astiz ME, Weil MH. Delayed lactate clearance in patients surviving circulatory shock. *Acute care.* 1985;11(3-4):212-215.
183. Jansen TC, van Bommel J, Woodward R, Mulder PG, Bakker J. Association between blood lactate levels, Sequential Organ Failure Assessment subscores, and 28-day mortality during early and late intensive care unit stay: a retrospective observational study. *Crit Care Med.* 2009;37(8):2369-2374.
184. Nguyen HB, Rivers EP, Knoblich BP, et al. Early lactate clearance is associated with improved outcome in severe sepsis and septic shock. *Crit Care Med.* 2004;32(8):1637-1642.
185. Nguyen HB, Kuan WS, Batech M, et al. Outcome effectiveness of the severe sepsis resuscitation bundle with addition of lactate clearance as a bundle item: a multi-national evaluation. *Crit Care.* 2011;15(5):R229.
186. Arnold RC, Shapiro NI, Jones AE, et al. Multicenter study of early lactate clearance as a determinant of survival in patients with presumed sepsis. *Shock.* 2009;32(1):35-39.
187. Drumheller BC, Agarwal A, Mikkelsen ME, et al. Risk factors for mortality despite early protocolized resuscitation for severe sepsis and septic shock in the emergency department. *J Crit Care.* 2016;31(1):13-20.
188. Chertoff J, Chisum M, Simmons L, King B, Walker M, Lascano J. Prognostic utility of plasma lactate measured between 24 and 48 h after initiation of early goal-directed therapy in the management of sepsis, severe sepsis, and septic shock. *J Intensive Care.* 2016;4:13.
189. Tian HH, Han SS, Lv CJ, et al. [The effect of early goal lactate clearance rate on the outcome of septic shock patients with severe pneumonia]. *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue.* 2012;24(1):42-45.
190. Gu WJ, Zhang Z, Bakker J. Early lactate clearance-guided therapy in patients with sepsis: a meta-analysis with trial sequential analysis of randomized controlled trials. *Intensive Care Med.* 2015;41(10):1862-1863.
191. Dellinger RP, Carlet JM, Masur H, et al. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Crit Care Med.* 2004;32(3):858-873.
192. Shankar-Hari M, Phillips GS, Levy ML, et al. Developing a New Definition and Assessing New Clinical Criteria for Septic Shock: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA.* 2016;315(8):775-787.
193. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med.* 2000;342(25):1887-1892.
194. Scheer CS, Fuchs C, Kuhn SO, et al. Quality Improvement Initiative for Severe Sepsis and Septic Shock Reduces 90-Day Mortality: A 7.5-Year Observational Study. *Crit Care Med.* 2017;45(2):241-252.
195. McColl T, Gatien M, Calder L, et al. Implementation of an Emergency Department Sepsis Bundle and System Redesign: A Process Improvement Initiative. *CJEM.* 2016:1-10.
196. Thompson MP, Reeves MJ, Bogan BL, DiGiovine B, Posa PJ, Watson SR. Protocol-Based Resuscitation Bundle to Improve Outcomes in Septic Shock Patients: Evaluation of the Michigan Health and Hospital Association Keystone Sepsis Collaborative. *Crit Care Med.* 2016.

197. Liu VX, Morehouse JW, Marelich GP, et al. Multicenter implementation of a treatment bundle for patients with sepsis and intermediate lactate values. *American journal of respiratory and critical care medicine.* 2016;193(11):1264-1270.
198. Aggarwal S, Noumi B, Velani S, Dasenbrook E, Attaway A. Improvements In Mortality Among Septic Patients Requiring Endotracheal Intubation: Cohort Study Of The Healthcare Cost And Utilization Project's Nationwide Inpatient Sample From 2001-2012. *Am J Resp Crit Care Med.* 2015;245(191):A1616.
199. Dombrovskiy VY, Martin AA, Sunderram J, Paz HL. Rapid increase in hospitalization and mortality rates for severe sepsis in the United States: A trend analysis from 1993 to 2003. *Crit Care Med.* 2007;35(5):1244-1250.
200. Ani C, Farshidpanah S, Bellinghausen Stewart A, Nguyen HB. Variations in organism-specific severe sepsis mortality in the United States: 1999-2008. *Crit Care Med.* 2015;43(1):65-77.
201. Stevenson EK, Rubenstein AR, Radin GT, Wiener RS, Walkey AJ. Two decades of mortality trends among patients with severe sepsis: a comparative meta-analysis. *Crit Care Med.* 2014;42(3):625-631.
202. Kumar G, Kumar N, Taneja A, et al. Nationwide trends of severe sepsis in the 21st century (2000-2007). *Chest.* 2011;140(5):1223-1231.
203. Attaway A, Aggarwal A, Noumi B, Velani S, Dasenbrook EC. Comparison of Mortality Among Septic Patients Requiring Endotracheal Intubation: Cohort Study of the Healthcare Cost and Utilization Project's Nationwide Inpatient Sample from 2001-2012. *Am J Resp Crit Care Med.* 2015;245(191):A1615.
204. Goto T, Yoshida K, Tsugawa Y, Filbin MR, Camargo CA, Hasegawa K. Mortality trends in U.S. adults with septic shock, 2005-2011: a serial cross-sectional analysis of nationally-representative data. *BMC infectious diseases.* 2016;16(1):1-8.
205. Stoller J, Halpin L, Weis M, et al. Epidemiology of severe sepsis: 2008-2012. *Journal of Critical Care.* 2016;31(1):58-62.
206. Chao-Ping W, Yu-Wei C, Vineet R, Ali H, Min-Jung W. Decreasing Trend of In-hospital Mortality for Sepsis, Severe Sepsis, and Septic Shock: A US Nationwide In-Patient Database Analysis from 2007-2011. *B47. CRITICAL CARE: CLINICAL RESEARCH DISCOVERIES, EPIDEMIOLOGY, AND HEALTH SERVICES FOR SEPSIS*: American Thoracic Society; 2016:A7824-A7824.
207. Kadri SS, Rhee C, Strich JR, et al. Estimating ten-year trends in septic shock incidence and mortality in united states academic medical centers using clinical data. *Chest.* 2016.
208. Lagu T, Rothberg MB, Shieh MS, Pekow PS, Steingrub JS, Lindenauer PK. Hospitalizations, costs, and outcomes of severe sepsis in the United States 2003 to 2007. *Crit Care Med.* 2012;40(3):754-761.
209. Padkin A, Goldfrad C, Brady AR, Young D, Black N, Rowan K. Epidemiology of severe sepsis occurring in the first 24 hrs in intensive care units in England, Wales, and Northern Ireland. *Crit Care Med.* 2003;31(9):2332-2338.
210. Gao F, Melody T, Daniels DF, Giles S, Fox S. The impact of compliance with 6-hour and 24-hour sepsis bundles on hospital mortality in patients with severe sepsis: a prospective observational study. *Crit Care.* 2005;9.
211. Reuben AD, Appelboam AV, Higginson I. The outcomes of severe sepsis and septic shock in the UK. *Crit Care.* 2006;10(4):417.
212. Melville J, Ranjan S, Morgan P. ICU mortality rates in patients with sepsis before and after the Surviving Sepsis Campaign. *Crit Care.* 2015;19(Suppl 1):P15-P15.
213. Daniels R. Surviving the first hours in sepsis: getting the basics right (an intensivist's perspective). *J Antimicrob Chemother.* 2011;66 Suppl 2:ii11-23.
214. Sivayoham N, Rhodes A, Jaiganesh T, van Zyl Smit N, Elkhodhair S, Krishnanandan S. Outcomes from implementing early goal-directed therapy for severe sepsis and septic shock : a 4-year observational cohort study. *Eur J Emerg Med.* 2012;19(4):235-240.
215. Mouncey PR, Osborn TM, Power GS, Harrison DA, Sadique MZ, Grieve RD. Trial of early, goal-directed resuscitation for septic shock. *N Engl J Med.* 2015;372.
216. Gray A, Ward K, Lees F, et al. The epidemiology of adults with severe sepsis and septic shock in Scottish emergency departments. *Emerg Med J.* 2013;30(5):397-401.
217. Finfer S, Bellomo R, Lipman J, French C, Dobb G, Myburgh J. Adult-population incidence of severe sepsis in Australian and New Zealand intensive care units. *Intensive Care Med.* 2004;30(4):589-596.
218. Kaukonen KM, Bailey M, Suzuki S, Pilcher D, Bellomo R. Mortality Related to Severe Sepsis and Septic Shock Among Critically Ill Patients in Australia and New Zealand, 2000-2012. *JAMA.* 2014;311(13):1308-1316.

219. Einsiedel LJ, Woodman RJ. Two nations: racial disparities in bloodstream infections recorded at Alice Springs Hospital, central Australia, 2001-2005. *Med J Aust.* 2010;192(10):567-571.
220. Burrell AR, McLaws ML, Fullick M, Sullivan RB, Sindhusake D. SEPSIS KILLS: early intervention saves lives. *Med J Aust.* 2016;204(2):73.
221. Investigators A, Group ACT, Peake SL, Delaney A, Bailey M, Bellomo R. Goal-directed resuscitation for patients with early septic shock. *N Engl J Med.* 2014;371.
222. Sundararajan V, Macisaac CM, Presneill JJ, Cade JF, Visvanathan K. Epidemiology of sepsis in Victoria, Australia. *Crit Care Med.* 2005;33(1):71-80.
223. Coz Yataco A, Jaehne AK, Rivers EP. Protocolized Early Sepsis Care Is Not Only Helpful for Patients: It Prevents Medical Errors. *Critical Care Medicine.* 2017;Publish Ahead of Print.
224. Nguyen HB, Jaehne AK, Jayaprakash N, et al. Early goal-directed therapy in severe sepsis and septic shock: insights and comparisons to ProCESS, ProMISe, and ARISE. *Critical Care.* 2016;20(1):1-16.
225. Castellanos-Ortega A, Suberviola B, Garcia-Astudillo LA, et al. Impact of the Surviving Sepsis Campaign protocols on hospital length of stay and mortality in septic shock patients: results of a three-year follow-up quasi-experimental study. *Crit Care Med.* 2010;38(4):1036-1043.
226. Laguna-Perez A, Chilet-Rosell E, Delgado Lacosta M, Alvarez-Dardet C, Uris Selles J, Munoz-Mendoza CL. Clinical pathway intervention compliance and effectiveness when used in the treatment of patients with severe sepsis and septic shock at an Intensive Care Unit in Spain. *Revista latino-americana de enfermagem.* 2012;20(4):635-643.
227. Memon JI, Rehmani RS, Alaithan AM, et al. Impact of 6-hour sepsis resuscitation bundle compliance on hospital mortality in a saudi hospital. *Crit Care Res Pract.* 2012;2012:273268.
228. Shiramizo SC, Marra AR, Durao MS, Paes AT, Edmond MB, Pavao dos Santos OF. Decreasing mortality in severe sepsis and septic shock patients by implementing a sepsis bundle in a hospital setting. *PLoS ONE.* 2011;6(11):e26790.
229. Apibunyopas Y. Mortality rate among patients with septic shock after implementation of 6-hour sepsis protocol in the emergency department of Thammasat University Hospital. *Journal Med Ass Thai.* 2014;97(Suppl 8):S182-193.
230. Cannon CM, Holthaus CV, Zubrow MT, et al. The GENESIS Project (GENeralized Early Sepsis Intervention Strategies): A Multicenter Quality Improvement Collaborative. *J Intensive Care Med.* 2013;28(6):787-795.
231. Casserly B, Baram M, Walsh P, Sucov A, Ward NS, Levy MM. Implementing a collaborative protocol in a sepsis intervention program: lessons learned. *Lung.* 2011;189(1):11-19.
232. Coba V, Whitmill M, Mooney R, et al. Resuscitation Bundle Compliance in Severe Sepsis and Septic Shock: Improves Survival, Is Better Late than Never. *J Intensive Care Med.* 2011;26(5):304-313.
233. Crowe CA, Kulstad EB, Mistry CD, Kulstad CE. Comparison of severity of illness scoring systems in the prediction of hospital mortality in severe sepsis and septic shock. *J Emerg Trauma Shock.* 2010;3(4):342-347.
234. Focht A, Jones AE, Lowe TJ. Early goal-directed therapy: improving mortality and morbidity of sepsis in the emergency department. *Jt Comm J Qual Patient Saf.* 2009;35(4):186-191.
235. Gatewood MOK, Wemple M, Greco S, Kritek PA, Durvasula R. A quality improvement project to improve early sepsis care in the emergency department. *BMJ quality & safety.* 2015(12):787-795.
236. Girardis M, Rinaldi L, Donno L, et al. Effects on management and outcome of severe sepsis and septic shock patients admitted to the intensive care unit after implementation of a sepsis program: a pilot study. *Crit Care.* 2009;13(5):R143.
237. Guerra WF, Mayfield TR, Meyers MS, Clouatre AE, Riccio JC. Early Detection And Treatment Of Patients With Severe Sepsis By Prehospital Personnel. *J Emerg Med.* 2013(6):1116-1125.
238. Guo Q, Li HY, Li YM, et al. Compliance with severe sepsis bundles and its effect on patient outcomes of severe community-acquired pneumonia in a limited resources country. *Archives of medical science : AMS.* 2014;10(5):970-978.
239. Gurnani PK, Patel GP, Crank CW, et al. Impact of the implementation of a sepsis protocol for the management of fluid-refractory septic shock: A single-center, before-and-after study. *Clin Ther.* 2010;32(7):1285-1293.
240. Jacob ST, Banura P, Baeten JM, et al. The impact of early monitored management on survival in hospitalized adult Ugandan patients with severe sepsis: A prospective intervention study. *Crit Care Med.* 2012;40(7):2050-2058.

241. Jones AE, Focht A, Horton JM, Kline JA. Prospective external validation of the clinical effectiveness of an emergency department-based early goal-directed therapy protocol for severe sepsis and septic shock. *Chest.* 2007;132(2):425-432.
242. Jones AE, Troyer JL, Kline JA. Cost-effectiveness of an emergency department-based early sepsis resuscitation protocol. *Crit Care Med.* 2011(6):1306-1312.
243. Kliger J, Singer SJ, Hoffman FH. Using the Integrated Nurse Leadership Program to Reduce Sepsis Mortality. *Jt Comm J Qual Patient Saf.* 2015;41(6):264-272.
244. Kuan WS, Mahadevan M, Tan JH, Guo J, Ibrahim I. Feasibility of introduction and implementation of the Surviving Sepsis Campaign bundle in a Singapore emergency department. *Eur J Emerg Med.* 2013;20(5):344-349.
245. Levy MM, Rhodes A, Phillips GS, et al. Surviving Sepsis Campaign: association between performance metrics and outcomes in a 7.5-year study. *Intensive Care Med.* 2014;40(11):1623-1633.
246. Miller RR, 3rd, Dong L, Nelson NC, et al. Multicenter implementation of a severe sepsis and septic shock treatment bundle. *Am J Respir Crit Care Med.* 2013;188(1):77-82.
247. Na S, Kuan WS, Mahadevan M, et al. Implementation of early goal-directed therapy and the surviving sepsis campaign resuscitation bundle in Asia. *International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua.* 2012;24(5):452-462.
248. Nguyen HB, Corbett SW, Steele R, Banta J, Clark RT, Hayes SR. Implementation of a bundle of quality indicators for the early management of severe sepsis and septic shock is associated with decreased mortality. *Crit Care Med.* 2007;35.
249. Noritomi DT, Ranzani OT, Monteiro MB, et al. Implementation of a multifaceted sepsis education program in an emerging country setting: clinical outcomes and cost-effectiveness in a long-term follow-up study. *Intensive Care Med.* 2014;40(2):182-191.
250. Qu HP, Qin S, Min D, Tang YQ. [The effects of earlier resuscitation on following therapeutic response in sepsis with hypoperfusion.]. *Zhonghua Wai Ke Za Zhi (Chinese Journal of Surgery).* 2006;44(17):1193-1196.
251. Rhodes A, Phillips G, Beale R, et al. The Surviving Sepsis Campaign bundles and outcome: results from the International Multicentre Prevalence Study on Sepsis (the IMPreSS study). *Intensive Care Med.* 2015;41(9):1620-1628.
252. Rinaldi L, Ferrari E, Marietta M, et al. Effectiveness of sepsis bundle application in cirrhotic patients with septic shock: a single-center experience. *J Crit Care.* 2013;28(2):152-157.
253. Salleh FM, Fathil SM, Zulkernain A, Zuraidah CM. Early goal-directed therapy in the management of severe sepsis/septic shock in an academic emergency department in Malaysia. *Crit Care & Shock.* 2010;13(3):91-97.
254. Schramm GE, Kashyap R, Mullon JJ, Gajic O, Afessa B. Septic shock: A multidisciplinary response team and weekly feedback to clinicians improve the process of care and mortality. *Crit Care Med.* 2011;39(2):252-258.
255. Siontis B, Elmer J, Dannielson R, et al. Multifaceted interventions to decrease mortality in patients with severe sepsis/septic shock—a quality improvement project. *PeerJ.* 2015;3:e1290.
256. Sohn CH, Ryoo SM, Seo DW, et al. Outcome of delayed resuscitation bundle achievement in emergency department patients with septic shock. *Intern Emerg Med.* 2014;9(6):671-676.
257. van Zanten ARH, Brinkman S, Arbous MS, et al. Guideline Bundles Adherence and Mortality in Severe Sepsis and Septic Shock. *Crit Care Med.* 2014;42(8):1890-1898.
258. Wang Z, Xiong Y, Schorr C, Dellinger RP. Impact of Sepsis Bundle Strategy on Outcomes of Patients Suffering from Severe Sepsis and Septic Shock in China. *J Emerg Med.* 2013(4):735-741.
259. Westphal GA, Koenig A, Caldeira Filho M, et al. Reduced mortality after the implementation of a protocol for the early detection of severe sepsis. *J Crit Care.* 2011;26(1):76-81.
260. Winterbottom F, Seoane L, Sundell E, Niazi J, Nash T. Improving Sepsis Outcomes for Acutely Ill Adults Using Interdisciplinary Order Sets. *Clin Nurse Spec.* 2011;25(4):180-185.
261. Yang M, Graham C, Rainer T. Outcome after implementation of sepsis guideline in the emergency department of a university hospital in Hong Kong. *Hong Kong J Emerg Med.* 2015;22(3):163-171.
262. Zambon M, Ceola M, Almeida-de-Castro R, Gullo A, Vincent JL. Implementation of the Surviving Sepsis Campaign guidelines for severe sepsis and septic shock: we could go faster. *J Crit Care.* 2008;23(4):455-460.
263. Azkarate I, Choperena G, Salas E, et al. Epidemiology and prognostic factors in severe sepsis/septic shock. Evolution over six years. *Med Intensiva.* 2015.

264. Worapratya P, Wanjaroenchaisuk A, Joraluck J, Wuthisuthimethawee P. Success of applying early goal-directed therapy for septic shock patients in the emergency department. *Open Access Emerg Med.* 2016;8:1-6.
265. Scheer CS, Fuchs C, Kuhn S-O, et al. Quality Improvement Initiative for Severe Sepsis and Septic Shock Reduces 90-Day Mortality: A 7.5-Year Observational Study. *Critical Care Medicine.* 2016;Publish Ahead of Print.
266. Leisman DE, Doerfler ME, Ward MF, et al. Survival Benefit and Cost Savings From Compliance With a Simplified 3-Hour Sepsis Bundle in a Series of Prospective, Multisite, Observational Cohorts. *Critical Care Medicine.* 2016;Publish Ahead of Print.
267. Leisman D, Wie B, Doerfler M, et al. Association of Fluid Resuscitation Initiation Within 30 Minutes of Severe Sepsis and Septic Shock Recognition With Reduced Mortality and Length of Stay. *Annals of Emergency Medicine.* 2016.
268. Assuncao MS, Teich V, Shiramizo SC, et al. The cost-effectiveness ratio of a managed protocol for severe sepsis. *J Crit Care.* 2014;29(4):692 e691-696.
269. El Solh AA, Akinnusi ME, Alsawalha LN, Pineda LA. Outcome of septic shock in older adults after implementation of the sepsis "bundle". *J Am Geriatr Soc.* 2008;56(2):272-278.
270. Lefrant JY, Muller L, Raillard A, et al. Reduction of the severe sepsis or septic shock associated mortality by reinforcement of the recommendations bundle: a multicenter study. *Ann Fr Anesth Reanim.* 2010;29(9):621-628.
271. MacRedmond R, Hollohan K, Stenstrom R, Nebre R, Jaswal D, Dodek P. Introduction of a comprehensive management protocol for severe sepsis is associated with sustained improvements in timeliness of care and survival. *Qual Saf Health Care.* 2010;19(5):e46.
272. Patel GW, Roderman N, Gehring H, Saad J, Bartek W. Assessing the effect of the Surviving Sepsis Campaign treatment guidelines on clinical outcomes in a community hospital. *Ann Pharmacother.* 2010;44(11):1733-1738.
273. Puskarich M, Marchick M, Kline J, Steuerwald M, Jones A. One year mortality of patients treated with an emergency department based early goal directed therapy protocol for severe sepsis and septic shock: A before and after study. *Crit Care.* 2009;13:R167.
274. Ryoo SM, Kim WY, Sohn CH, et al. Prognostic value of timing of antibiotic administration in patients with septic shock treated with early quantitative resuscitation. *Am J Med Sci.* 2015;349(4):328-333.
275. Sebat F, Johnson D, Musthafa AA, et al. A multidisciplinary community hospital program for early and rapid resuscitation of shock in nontrauma patients. *Chest.* 2005;127(5):1729-1743.
276. Shapiro NI, Howell MD, Talmor D, et al. Implementation and outcomes of the Multiple Urgent Sepsis Therapies (MUST) protocol. *Crit Care Med.* 2006;34(4):1025-1032.
277. Hanzelka KM, Yeung SC, Chisholm G, et al. Implementation of modified early-goal directed therapy for sepsis in the emergency center of a comprehensive cancer center. *Support Care Cancer.* 2013;21(3):727-734.
278. Jeon K, Shin TG, Sim MS, et al. Improvements in compliance with resuscitation bundles and achievement of end points after an educational program on the management of severe sepsis and septic shock. *Shock.* 2012;37(5):463-467.
279. Kang MJ, Shin TG, Jo IJ, et al. Factors influencing compliance with early resuscitation bundle in the management of severe sepsis and septic shock. *Shock.* 2012;38(5):474-479.
280. Kortgen A, Niederprum P, Bauer M. Implementation of an evidence-based "standard operating procedure" and outcome in septic shock. *Crit Care Med.* 2006;34(4):943-949.
281. Mikkelsen ME, Gaieski DF, Goyal M, et al. Factors associated with nonadherence to early goal-directed therapy in the ED. *Chest.* 2010;138(3):551-558.
282. Thiel SW, Asghar MF, Micek ST, Reichley RM, Doherty JA, Kollef MH. Hospital-wide impact of a standardized order set for the management of bacteremic severe sepsis. *Crit Care Med.* 2009;37(3):819-824.
283. Tromp M, Bleeker-Rovers CP, van Achterberg T, Kullberg BJ, Hulscher M, Pickkers P. Internal medicine residents' knowledge about sepsis: effects of a teaching intervention. *Neth J Med.* 2009;67(9):312-315.
284. Trzeciak S, Dellinger RP, Abate NL, et al. Translating research to clinical practice: a 1-year experience with implementing early goal-directed therapy for septic shock in the emergency department. *Chest.* 2006;129(2):225-232.
285. Wawrzeniak I, Loss S, Moraes M, De La Vega F, Victorino J. Could a protocol based on early goal-directed therapy improve outcomes in patients with severe sepsis and septic shock in the Intensive Care Unit setting? *Indian J Crit Care Med.* 2015;19(3):159-165.

286. Carvas JM, Canelas C, Montanha G, Silva C, Esteves F. Impact of Compliance with a Sepsis Resuscitation Bundle in a Portuguese Emergency Department. *Acta Med Port.* 2016;29(2):88-94.
287. Early Goal-Directed Therapy Collaborative Group of Zhejiang P. [The effect of early goal-directed therapy on treatment of critical patients with severe sepsis/septic shock: a multi-center, prospective, randomized, controlled study]. *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue.* 2010;22(6):331-334.
288. Andrews B, Muchemwa L, Kelly P, Lakhi S, Heimburger DC, Bernard GR. Simplified severe sepsis protocol: a randomized controlled trial of modified early goal-directed therapy in Zambia. *Crit Care Med.* 2014;42(11):2315-2324.
289. Chen ZQ, Jin YH, Chen H, Fu WJ, Yang H, Wang RT. [Early goal-directed therapy lowers the incidence, severity and mortality of multiple organ dysfunction syndrome]. *Nan Fang Yi Ke Da Xue Xue Bao.* 2007;27(12):1892-1895.
290. Hayes MA, Timmins AC, Yau EH, Palazzo M, Hinds CJ, Watson D. Elevation of systemic oxygen delivery in the treatment of critically ill patients. *N Engl J Med.* 1994;330(24):1717-1722.
291. He ZY, Gao Y, Wang XR, Hang YN. [Clinical evaluation of execution of early goal directed therapy in septic shock]. *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue.* 2007;19(1):14-16.
292. Mouncey PR, Osborn TM, Power GS, et al. Trial of early, goal-directed resuscitation for septic shock. *N Engl J Med.* 2015;372(14):1301-1311.
293. Investigators A, Group ACT, Peake SL, et al. Goal-directed resuscitation for patients with early septic shock. *N Engl J Med.* 2014;371(16):1496-1506.
294. Tuchschmidt J, Fried J, Astiz M, Rackow E. Elevation of cardiac output and oxygen delivery improves outcome in septic shock. *Chest.* 1992;102(1):216-220.
295. Wang XZ, Lu CJ, Gao FQ, Li XH, Yan WF, Ning FY. [Efficacy of goal-directed therapy in the treatment of septic shock]. *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue.* 2006;18(11):661-664.
296. Yan J, Cai G. A multicentre study on early goal-directed therapy of severe sepsis and septic shock patients in the ICU: collaborative study group on early goal-directed therapy in Zhejiang Province, China. *Crit Care.* 2008;12(Suppl 2):P417.
297. ProCESS I, Yealy DM, Kellum JA, et al. A randomized trial of protocol-based care for early septic shock. *N Engl J Med.* 2014;370(18):1683-1693.
298. Yu M, Burchell S, Hasaniya NW, Takanishi DM, Myers SA, Takiguchi SA. Relationship of mortality to increasing oxygen delivery in patients > or = 50 years of age: a prospective, randomized trial. *Crit Care Med.* 1998;26(6):1011-1019.
299. Yu M, Levy MM, Smith P, Takiguchi SA, Miyasaki A, Myers SA. Effect of maximizing oxygen delivery on morbidity and mortality rates in critically ill patients: a prospective, randomized, controlled study. *Crit Care Med.* 1993;21(6):830-838.
300. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med.* 2001;345(19):1368-1377.
301. Jaehne AK, Rivers EP. Early Liberal Fluid Therapy for Sepsis Patients Is Not Harmful: Hydrophobia Is Unwarranted but Drink Responsibly. *Critical Care Medicine.* 2016;44(12):2263-2269.
302. Wira CR, Dodge K, Sather J, Dziura J. Meta-analysis of Protocolized Goal-Directed Hemodynamic Optimization for the Management of Severe Sepsis and Septic Shock in the Emergency Department. *West J Emerg Med.* 2014;15(1):51-59.
303. Jones AE, Brown MD, Trzeciak S, et al. The effect of a quantitative resuscitation strategy on mortality in patients with sepsis: a meta-analysis. *Crit Care Med.* 2008;36(10):2734-2739.
304. Wang AT, Liu F, Zhu X, Yao GQ. [The effect of an optimized resuscitation strategy on prognosis of patients with septic shock: a systematic review]. *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue.* 2012;24(1):13-17.
305. Damiani E, Donati A, Serafini G, et al. Effect of performance improvement programs on compliance with sepsis bundles and mortality: a systematic review and meta-analysis of observational studies. *PLoS One.* 2015;10(5):e0125827.
306. Gu WJ, Wang F, Bakker J, Tang L, Liu JC. The effect of goal-directed therapy on mortality in patients with sepsis - earlier is better: a meta-analysis of randomized controlled trials. *Crit Care.* 2014;18(5):570.

307. Chelkeba L, Ahmadi A, Abdollahi M, Najafi A, Mojtahedzadeh M. Early goal-directed therapy reduces mortality in adult patients with severe sepsis and septic shock: Systematic review and meta-analysis. *Indian J Crit Care Med.* 2015;19(7):401-411.
308. Rusconi AM, Bossi I, Lampard JG, Szava-Kovats M, Bellone A, Lang E. Early goal-directed therapy vs usual care in the treatment of severe sepsis and septic shock: a systematic review and meta-analysis. *Intern Emerg Med.* 2015;10(6):731-743.
309. Lu J, Wang X, Chen Q, et al. The effect of early goal-directed therapy on mortality in patients with severe sepsis and septic shock: a meta-analysis. *Journal of Surgical Research.*
310. Yu H, Chi D, Wang S, Liu B. Effect of early goal-directed therapy on mortality in patients with severe sepsis or septic shock: a meta-analysis of randomised controlled trials. *BMJ Open.* 2016;6(3).
311. Jiang L-b, Zhang M, Jiang S-y, MA Y-f. Early goal-directed resuscitation for patients with severe sepsis and septic shock: a meta-analysis and trial sequential analysis. *Scandinavian journal of trauma, resuscitation and emergency medicine.* 2016;24(1):1-12.
312. Angus DC, Barnato AE, Bell D, et al. A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, ProCESS and ProMISe Investigators. *Intensive Care Med.* 2015;41(9):1549-1560.
313. Xing L, Tong L, Jun L, Xinjing G, Lei X. [Effect of early goal-directed therapy on mortality in patients with severe sepsis or septic shock: a Meta analysis]. *Zhonghua wei zhong bing ji jiu yi xue.* 2015;27(9):735-738.
314. Zhang L, Zhu G, Han L, Fu P. Early goal-directed therapy in the management of severe sepsis or septic shock in adults: a meta-analysis of randomized controlled trials. *BMC Medicine.* 2015;13(1):1-12.
315. Simpson SQ, Gaines M, Hussein Y, Badgett RG. Early goal-directed therapy for severe sepsis and septic shock: a living systematic review. *Journal of Critical Care.* 2016.
316. Lee WK, Kim HY, Lee J, Koh SO, Kim JM, Na S. Protocol-Based Resuscitation for Septic Shock: A Meta-Analysis of Randomized Trials and Observational Studies. *Yonsei medical journal.* 2016;57(5):1260-1270.
317. Xu JY, Chen QH, Liu SQ, et al. The Effect of Early Goal-Directed Therapy on Outcome in Adult Severe Sepsis and Septic Shock Patients: A Meta-Analysis of Randomized Clinical Trials. *Anesth Analg.* 2016.
318. Coccolini F, Sartelli M, Catena F, Ceresoli M, Montori G, Ansaloni L. Early goal-directed treatment versus standard care in management of early septic shock: Meta-analysis of randomized trials. *The journal of trauma and acute care surgery.* 2016.
319. Liu B, Ding X, Yang J. Effect of early goal directed therapy in the treatment of severe sepsis and/or septic shock. *Current medical research and opinion.* 2016:1-10.
320. Guolong C, Hongjie T, Xuejing H, et al. The effects of early goal-directed therapy on mortality rate in patients with severe sepsis and septic shock: A systematic literature review and meta-analysis. *Chinese Critical Care Medicine.* 2016;28(6):439-442.
321. Winters ME, Sherwin R, Vilke GM, Wardi G. Does Early Goal-Directed Therapy Decrease Mortality Compared with Standard Care in Patients with Septic Shock? *J Emerg Med.* 2016.

THIS FORM HAS BEEN UPDATED SINCE DATE OF SUBMISSION

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
0500_Evidence_CompositeMSF1_0_Data_Form-2_3_2017.doc
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.
The evidence cited for all components of this measure is directly related to decreases in organ failure, overall reductions in hospital mortality, length of stay, and costs of care.

A principle of sepsis care is that clinicians must rapidly treat patients with an unknown causative organism and unknown antibiotic susceptibility. Since patients with severe sepsis have little margin for error regarding antimicrobial therapy, initial treatment should be broad spectrum to cover all likely pathogens. As soon as the causative organism is identified, based on subsequent culture and susceptibility testing, de-escalation is encouraged by selecting the most appropriate antimicrobial therapy to cover the identified pathogen, safely and cost effectively (Dellinger, 2012).

Multicenter efforts to promote bundles of care for severe sepsis and septic shock were associated with improved guideline compliance and lower hospital mortality (Ferrer, 2008 and Rhodes, 2015). Even with compliance rates of less than 30%, absolute reductions in mortality of 4-6% have been noted (Levy, 2010 and Ferrer, 2008). Absolute reductions in mortality of over 20% have been seen with compliance rates of 52% (Levy, 2010). Coba et al. has shown that when all bundle elements are completed and compared to patients who do not have bundle completion, the mortality difference is 14% (2011). Thus, there is a direct association between bundle compliance and improved mortality. Without a continuous quality initiative (CQI), even these compliance rates will not improve and will decrease over time (Ferrer, 2008). Multiple studies have shown that, for patients with severe sepsis, standardized order sets, enhanced bedside monitor display, telemedicine, and comprehensive CQI feedback is feasible, modifies clinician behavior, and is associated with decreased hospital mortality (Thiel, 2009; Micek, 2006; Winterbottom, 2011; Schramm, 2011; Nguyen, 2007; Loyola, 2011).

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 refer to the measure performance table in the separate table appendix attachment.

Below we include distribution information on performance rates calculated for three quarters, between Q4 2015 (the first year CMS

included this measure as part of the Hospital IQR program) and Q2 2016. Using three quarters of data, overall performance (across all hospitals) ranged from 34.40% to 44% and increasing over time.

In each quarter there is a wide range in hospital performance, indicating a continued need to implement the measure. Additionally, overall performance is low signaling a continued room for improvement in severe sepsis and septic shock care nationwide.

Q4 2015 Analysis Provider Level
3,134 hospitals submitted 96,516 eligible cases
Overall performance rate 34.40%
Min 0
10th percentile 5.0%
25th percentile 17.9%
Median 31.0%
75th percentile 45.8%
90th percentile 60.0%
Max 100.0%
Average 32.60%
Standard Deviation 21.10%

Q1 2016 Analysis Provider Level
3,182 hospitals submitted 104,166 eligible cases.
Overall performance rate 39.50%
Min 0
10th percentile 7.7%
25th percentile 21.6%
Median 36.1%
75th percentile 51.3%
90th percentile 66.7%
Max 100
Average 37.10%
Standard Deviation 21.90%

Q2 2016 Analysis Provider Level
3,193 hospitals submitted 101,599 eligible cases.
Overall performance rate 44.00%
Min 0
10th percentile 12.5%
25th percentile 25.8%
Median 41.7%
75th percentile 57.1%
90th percentile 71.4%
Max 100
Average 41.90%
Standard Deviation 22.90%

The chart below describes the categorization of 3 and 6 hour elements. For purposes of this analysis the 30 ml/kg crystalloid fluid started was grouped with the shock bundles, although technically only refractive hypotension qualifies for shock.

<table>
<tr><th rowspan="2">Required Action</th><th colspan="2">Severe Sepsis</th><th colspan="2">Septic Shock</th></tr>
<tr><th>3 hour Bundle</th><th>6 hour Bundle</th><th>3 hour Bundle</th><th>6 hour Bundle</th></tr>
<tr><td>Initial Lactate Collection</td><td>Yes</td><td colspan="3" rowspan="3">Must be completed within 3 hours of Severe Sepsis Presentation</td></tr>
<tr><td>Blood Culture Collection</td><td>Yes</td></tr>
<tr><td>Initial Antibiotic Started</td><td>Yes</td></tr>
<tr><td>Repeat Lactate Collection (if Initial Lactate > 2)</td><td colspan="2">Yes</td><td colspan="2">Must be completed within 6 hours of Severe Sepsis presentation</td></tr>
<tr><td>30mL/kg Crystalloid Fluids Started</td><td>N/A</td><td>N/A</td><td>Yes</td><td>Must be completed within 3 hrs of Septic Shock</td></tr>
<tr><td>Vasopressor Given (if ↓ BP persists)</td><td>N/A</td><td>N/A</td><td rowspan="2">Must be completed within 6 hrs of Septic Shock</td><td>Yes</td></tr>
<tr><td>Repeat Volume Status/ Tissue Perfusion Assessment</td><td>N/A</td><td>N/A</td><td>Yes</td></tr>
</table>

Below are the performance rates for each component broken down by 3 and 6 hour bundle groupings. Significant gaps are demonstrated at each level of performance when all charts are analyzed for Q3 2015 through Q2 2016 in the chart below. The repeat volume and perfursion assessment data below is broken down into focused exam and hemodynamic elements that are no longer required. No data is yet available on the new attestation strategy.

Population Description	Cases	Bundle Percentage
Initial Population Number of Sepsis Cases	**325,809**	**--**
Total Number of Excluded Sepsis Cases	*166,520*	--
Total Number of Eligible Sepsis Cases	**159,289**	**--**
Total Number of Passed Sepsis Cases	*64,051*	--
Total Number of Failed Sepsis Cases	*95,238*	--
Severe Sepsis 3 Hour Bundle Eligible Cases	**167,114**	**--**
Severe Sepsis 3 Hour Bundle Passes	*110,078*	*65.9%*
Severe Sepsis 3 Hour Bundle Failures	*54,618*	*32.7%*
Initial Lactate Level Failures	26,503	48.5%
Broad Spectrum or Other Antibiotic Administration Failures	20,951	38.4%
Blood Culture Collection Failures	18,772	34.4%
Severe Sepsis 6 Hour Bundle Eligible Cases	**90,385**	**--**
Severe Sepsis 6 Hour Bundle Passes	*53,475*	*59.2%*
Severe Sepsis 6 Hour Bundle Failures	*36,910*	*40.8%*
Repeat Lactate Level Failures	36,910	

Septic Shock 3 Hour Bundle Eligible Cases	**40,989**	--
Septic Shock 3 Hour Bundle Passes	*22,359*	*54.5%*
Septic Shock 3 Hour Bundle Failures	*18,630*	*45.5%*
Crystalloid Fluid Administration Failures	18,630	
Vasopressor Shock 6 Hour Bundle Eligible Cases	**8,177**	--
Vasopressor Shock 6 Hour Bundle Passes	*6,157*	*75.3%*
Vasopressor Shock 6 Hour Bundle Failures	*2,020*	*24.7%*
Vasopressor Administration Failures	2,020	
Focus Exam Shock 6 Hour Bundle Eligible Cases	**14,630**	--
Focus Exam Shock 6 Hour Bundle Passes	*3,801*	*26.0%*
Focus Exam Shock 6 Hour Bundle Failures	*9,935*	*67.9%*
Hemodynamic Choices Shock 6 Hour Bundle Eligible Cases	**10,829**	--
Hemodynamic Choices Shock 6 Hour Bundle Passes	*894*	*8.3%*
Hemodynamic Choices Shock 6 Hour Bundle Failures	*9,935*	*91.7%*

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.*
Results of our disparity analysis are summarized below. Please refer to the sepsis disparities table in the table appendix attachment for additional details.

Across every age group, the performance rate (receipt of all applicable elements of care) increases with each successive quarter (for example, 18-49: 34.36%, 39.60%, 43.59%).

Within each quarter, the performance across the different age categories tends to be similar (18-49: 34.36%, 50-64: 33.47, 65-84: 35.01, 85+: 34.13).

When comparing the Hispanic group to the non-Hispanic group, the non-Hispanic group consistently has a higher proportion of patients receiving all applicable elements of clinical care across all the quarters (2015 Q4: 34.53% vs 32.64%, 2016 Q1: 39.80% vs 36.35%, 2016 Q2: 44.23% vs 40.82%). These results were statistically significant (p-value < 0.05).

Across genders, females and males have a similar proportion of patients receiving all applicable care elements (2015 Q4: 33.82% vs 34.93%, 2016 Q1: 39.22% vs 39.86%, 2016 Q2: 43.28% vs 44.63%).

When comparing Medicare patients to non-Medicare patients, Medicare patients received all applicable elements of care at a higher proportion compared to non-Medicare patients consistently across all the quarters (2015 Q4: 34.90% vs 33.32%, 2016 Q1: 40.07% vs 38.47%, 2016 Q2: 44.57% vs 42.76%). These results were statistically significant (p-value < 0.05).

Race

When comparing performance rates across different races, there is a consistently lower proportion of blacks compared to whites and other races who receive all the elements of applicable care. Compared to whites, patients from other races had a similar rate of patients receiving all applicable elements of care across all quarters(Q4 '15, Q1' 16, Q2' 16) of submitted data (Black: 30.64% 35.93% 40.29%; White: 34.95% 40.08% 44.58%; Other: 34.95% 40.01% 43.82%). These results were statistically significant (p-value < 0.05).

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

1c. Composite Quality Construct and Rationale

1c.1. **A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.**

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
    - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient);

1c.1. Please identify the composite measure construction: all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)

1c.2. **Describe the quality construct, including:**

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

The overall area of quality under consideration is care of patients with severe sepsis or septic shock. The components are clearly articulated below in the numerator statement and include lactate collection, delivery of broad spectrum antibiotics, obtaining blood cultures, delivering resuscitation fluids, applying vasopressors as needed, reassessing volume and perfusion status and repeating lactate values. The relationship of the component measures to the overall composite is such that all components for which the individual case is eligible must be met or the individual case fails.

1c.3. **Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.**

A composite measure was developed given the dependencies the components have on one another. In addition, the components of the measure must be applied within specific time frames. The sequencing of the measure is such that the components could not stand alone unless certain preceding conditions had been met. In this way, treating the elements as a composite ensured assessment of a concerted strategy aimed at reducing mortality. The composite is more powerful that any individual application of the components in isolation from each other.

1c.4. **Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.**

The component measures are aggregated both in time with 3 and 6 hour elements for severe sepsis and for septic shock. In addition to being time based, proceeding with the next component depends on certain qualifying features creating dependencies within the measure. There is no weighting of one element as more important than another. This structure is consistent with the stated quality construct of providing measurement an orderly standard operating procedure in the management of patients with severe sepsis and septic shock.

THIS FORM HAS BEEN UPDATED SINCE DATE OF SUBMISSION

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):*
Critical Care, Infectious Diseases (ID), Infectious Diseases (ID) : Pneumonia and respiratory infections, Respiratory, Respiratory : Pneumonia

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):*
Elderly

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.)*
https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1228775749207

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:** Copy_of_Appendix_A.1_v5_2a-1-.xls

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.
Yes

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.
Since last submission in November 2014, Center for Medicare & Medicaid (CMS) adopted the measure for the Hospital Inpatient Quality Reporting (IQR) Program. The measure developer and leads, in collaboration with CMS, have continued to refine the specifications and data element definitions to increase clarity and reduce burden for abstractors. The measure has undergone at least three rounds of updates since the last endorsement, in alignment with published IQR Specification Manual and data dictionary, beginning with initial publication in Manual v5.0b and updates in v5.1, v5.2a, and v5.3.

v5.0b is relevant to patients who were discharged from the hospital from 10/2015 to 6/2016. v5.1 updates are relevant to patients who were discharged from 7/2016 to 12/2016. v5.2a updates are relevant to patients who were/will be discharged between 1/2017 to 12/2017. v5.3 updates will be relevant to patients who will be discharged between 1/2018 to 12/2018. Below we describe the changes in alphabetical order.

• Administrative Contraindication to Care, Septic Shock – New data element in v5.1

Change and reason: Previously, patient refusal of care at any point during the stay would result in excluding a case. This refusal required searching the entire medical record from arrival to discharge for documentation of patient refusal. The changes limit the

timeframe within which a patient refusal must be documented in order to exclude a patient; this coincides with the time-frames within which elements of care required to meet the measure must occur. This change reduces abstraction burden and makes the refusal timeframe more appropriate for the care being measured.

• Change in Abstraction requirement starting in v5.1

Change and reason: Previously when a case failed at any point in the measure, abstractors were obligated to continue collecting data. Revisions no longer require continued abstraction of additional data once a measure failure point is reached, which reduces abstraction burden.

• Blood Culture Collection Acceptable Delay – New data element in v5.2a

Reason: Previously, if a blood culture was not collected before IV antibiotics were administered, the case would fail. There are some situations where not collecting a blood culture before IV antibiotics are administered is clinically acceptable. This new data element captures instances where there was an acceptable delay in the collection of a blood culture and allows those cases to meet this part of the measure.

• Broad Spectrum or Other Antibiotic Administration Selection – Data element updated in v5.1 and v5.2a

Change and reason for v5.1: Previously, clinicians were required to order IV antibiotics as specified by tables which were part of the measure. This requirement limited a clinician's ability to order more focused treatment based on their clinical impressions in cases where the causative organism and susceptibility was known. The requirements of the data elements were expanded to allow for an exception in cases where the causative organism and susceptibility of that organism to a particular antibiotic is known.
Change and reason for v5.2a: The measure was updated to give clinicians more latitude in cases where C. difficile is identified as the causative organism.

• Clinical Trial – New data element for v5.3

Reason: Patients enrolled in clinical trials related to sepsis management are kept in the measure. These cases have an increased chance of not meeting the elements of care as prescribed by the measure due to the experimental nature of some treatment. This data element is being added so patients enrolled in clinical trials are excluded from the measure.

• Crystalloid Fluid Administration - Data element updated in v5.1 and v5.2a

Change and reason: List of acceptable crystalloid fluids was expanded from normal saline and Lactated Ringers solution to include balanced crystalloid solutions such as Normosol and Plasmalyte based on new literature and requests from facilities using these fluids. List of electrolyte additives to normal saline considered acceptable was added. This change allows volume of normal saline to which electrolytes have been added to be counted toward the 30 mL/kg target volume.

Allowance for ordered volumes up to 10% less than 30 mL/kg target volume was added. This accounts for variations in ordered volume in situations that require urgent administration of fluids and the patient weight is unknown.

Terms more consistent with actual ordering practices identified by many facilities were added as acceptable.
Revisions that allow patients with implanted ventricular assist devices (VAD) to bypass the requirement for 30 mL/kg of crystalloid fluids and subsequent data elements in the algorithm flow were made. This change keeps these patients in the measure for evaluation of elements of care for severe sepsis care.

• Change in abstraction requirement in v5.2a

Change and reason: Previously, data abstractors were not allowed to use documentation in pre-hospital records that were a part of the patient's medical record. This requirement was causing cases being abstracted to fail and/or miss out on important documentation. As a result, the measure was updated to allow documentation in pre-hospital records as long as it was a part of the medical record.

• Volume and perfusion reassessment requirement - Change in abstraction requirement in v5.2a and in v5.3

V5.2a Change and reason: Previously, documentation of any one of the five focused exam data elements being completed without

clinician documented results was not acceptable. To decrease clinician documentation burden and abstractor burden, the measure was updated to accept documentation that a clinician performed or attested to performing a physical exam, perfusion exam or sepsis focused exam to satisfy the requirements for each of the five focused exam data elements.

V5.3 Change and reason: Recent trials of early protocolized care in sepsis have shown similar mortality reductions with a variety of methods of volume and perfusion assessments, e.g. "usual care," which may be used as alternatives to quantitative resuscitation. Given these findings, there is little consensus on the best methods of reassessment, however there is consensus that reassessment itself is a best practice.1 As such, the measure no longer requires a provider to state the method of reassessment used (e.g. physical exam, bedside cardiovascular ultrasound, passive leg raising, CVP, ScV02 assessment). The clinician can attest that volume and perfusion reassessment has occurred, even without reference to the method used. This will meet the measure's volume and perfusion reassessment requirement. A provider may also opt to state their chosen method, but this is not required.

• Documentation of Septic Shock – New data element in v5.1

Reason: This data element was added as trigger point in the algorithm for administration of crystalloid fluids. This helps better identify which patients should receive crystalloid fluids.

• Initial Hypotension – New data element in v5.1

Reason: This data element was added as trigger point in the algorithm for administration of crystalloid fluids. This update helps better identify which patients should receive crystalloid fluids.

• Septic Shock present & Severe Sepsis present - Change in abstraction requirement in v5.2a

Change and reason: Previously, all physiological criteria, which represented SIRS criteria or organ dysfunction, would be used to determine presence of severe sepsis and/or septic shock. In some cases, abnormal values may be normal for a patient due to a chronic condition, a medication or something other than an infection. Changes were made to allow for abnormal values to be disregarded if sufficient and appropriate clinician documentation is provided.

• Vasopressor Administration - Change in abstraction requirement in v5.2a and v5.3

Change and reason in v5.2a: Previously, only intravenous (IV) vasopressor administration was considered an acceptable route. In some cases obtaining IV access is limited. The measure was updated to allow for an additional route, intraosseous vasopressor administration as well as IV.

Change and reason in v5.3: Previously, there were no allowances for cases where additional crystalloid fluids were given instead of vasopressors. This will be accounted for in v5.3.

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).

The number of patients in the denominator who received ALL of the following components (if applicable) for the early management of severe sepsis and septic shock: initial lactate levels, blood cultures, antibiotics, fluid resuscitation, repeat lactate level, vasopressors, and volume status and tissue perfusion reassessment.

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).
In addition to the previous information (above) about assessing the numerator population, the following also are part of the numerator

details.

- Within 3 hours of presentation of severe sepsis:
  - Measure initial lactate level
  - Draw blood cultures prior to antibiotics
  - Administer broad spectrum or other antibiotics
- Within 6 hours of presentation of severe sepsis:
  - Repeat lactate level (if initial lactate > 2 mmol/L)
- Within 3 hours of presentation of septic shock:
  - Administer 30 ml/kg crystalloid for hypotension or lactate ≥ 4 mmol/L
- Within 6 hours of presentation of septic shock:
  - Apply vasopressors (for hypotension that does not respond to initial fluid resuscitation) to maintain a mean arterial pressure (MAP) ≥ 65 mm Hg
  - Reasses volume status and tissue perfusion in the event of persistent hypotension (MAP <65 mm Hg) after initial fluid administration or initial lactate level ≥ 4 mmol/L
    - The clinician is no longer required to state the method of reassessment used (e.g. physical exam, bedside cardiovascular ultrasound, passive leg raising, CVP, ScV02 assessment). The clinician can attest that volume and perfusion reassessment has occurred, even without reference to the method used. This will meet the measure's volume and perfusion reassessment requirement. A provider may also opt to state their chosen method, but this is not required.

The full definitions of each of these data elements and additional information are included in the data dictionary, attached to the submission

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Inpatients age 18 and over with an ICD-10-CM Principal or Other Diagnosis Code of Sepsis, Severe Sepsis, or Septic Shock.

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).
Discharges age 18 and over with an ICD-10-CM Principal or Other Diagnosis Code of Sepsis, Severe Sepsis, or Septic Shock as defined in the table below:

ICD-10-CM Code	Code Description
A021	Salmonella sepsis
A227	Anthrax sepsis
A267	Erysipelothrix sepsis
A327	Listerial sepsis
A400	Sepsis due to streptococcus, group A
A401	Sepsis due to streptococcus, group B
A403	Sepsis due to Streptococcus pneumoniae
A408	Other streptococcal sepsis
A409	Streptococcal sepsis, unspecified
A4101	Sepsis due to Methicillin susceptible Staphylococcus aureus
A4102	Sepsis due to Methicillin resistant Staphylococcus aureus
A411	Sepsis due to other specified staphylococcus
A412	Sepsis due to unspecified staphylococcus
A413	Sepsis due to Hemophilus influenzae
A414	Sepsis due to anaerobes
A4150	Gram-negative sepsis, unspecified
A4151	Sepsis due to Escherichia coli [E. coli]
A4152	Sepsis due to Pseudomonas
A4153	Sepsis due to Serratia

A4159 Other Gram-negative sepsis
A4181 Sepsis due to Enterococcus
A4189 Other specified sepsis
A419 Sepsis, unspecified organism
A427 Actinomycotic sepsis
A5486 Gonococcal sepsis
B377 Candidal sepsis
R6520 Severe sepsis without septic shock
R6521 Severe sepsis with septic shock

Data elements required to calculate the denominator (in alphabetical order):

- Administrative Contraindication to Care, Septic Shock
- Administrative Contraindication to Care, Severe Sepsis
- Admission Date
- Birthdate
- Directive for Comfort Care or Palliative Care, Septic Shock
- Directive for Comfort Care or Palliative Care, Severe Sepsis
- Discharge Date
- Discharge Disposition
- Discharge Time
- Transfer From Another Hospital or ASC

The full definitions of each of these data elements and additional information are included in the data dictionary, attached to the submission

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The following patients are excluded from the denominator:

- Severe sepsis is not present
- Patients Transferred in from another acute care facility
- Patients receiving IV antibiotics for more than 24 hours prior to presentation of severe sepsis.
- Patients with a Directive for Comfort Care or Palliative Care within 3 hours of presentation of severe sepsis
- Patients with an Administrative Contraindication to Care within 6 hours of presentation of severe sepsis
- Patients with an Administrative Contraindication to Care within 6 hours of presentation of septic shock
- Patients with a Directive for Comfort Care or Palliative Care within 6 hours of presentation of septic shock
- Patients with septic shock who are discharged within 6 hours of presentation
- Patients with severe sepsis who are discharged within 6 hours of presentation
- Patients with a Length of Stay >120 days
- Patients included in a Clinical Trial

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.)*
To determine the length of stay, the admission date and discharge date are entered. If the result of the calculation subtracting the admission date from the discharge date is greater than 120 days, the patient is excluded from the measure.

Data Elements required to determine denominator exclusions (in alphabetical order):
Administrative Contraindication to Care, Septic Shock
Administrative Contraindication to Care, Severe Sepsis
Admission Date
Birthdate
Clinical Trial
Directive for Comfort Care or Palliative Care, Septic Shock
Directive for Comfort Care or Palliative Care, Severe Sepsis
Discharge Date
Discharge Disposition
Discharge Time

Transfer from Another Hospital or ASC

The full definitions of each of these data elements and additional information are included in the data dictionary, attached to the submission.

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.)*
N/A. This measure is not stratified.

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.)*
The detailed measure algorithm for SEP-1 is available in the data dictionary attached to the submission, along with a diagram.

1. Find the patients who meet the initial population (ie, the general group of patients that a set of performance measures is designed to address).
2. From the patients within the initial population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria) Note: in some cases the initial population and denominator are identical. Remove any patients that meet the denominator exclusion criteria.
3. The following actions are required within 3 hours of presentation of severe sepsis:
    a. Measure initial lactate level
    b. Draw blood cultures prior to antibiotics
    c. Administer broad spectrum or other antibiotics
4. The following actions are required within 6 hours of presentation of severe sepsis:
    a. Repeat lactate level (if initial lactate > 2 mmol/L)
5. The following actions are required within 3 hours of presentation of septic shock:
    a. Administer 30 ml/kg crystalloid for hypotension or lactate ≥ 4 mmol/L
6. The following actions are required within 6 hours of presentation of septic shock:
    a. Apply vasopressors (for hypotension that does not respond to initial fluid resuscitation) to maintain a mean arterial pressure (MAP) ≥ 65 mm Hg
    b. Reasses volume status and tissue perfusion in the event of persistent hypotension (MAP <65 mm Hg) after initial fluid administration or initial lactate level ≥ 4 mmol/L
    Note: The clinician is no longer required to state the method of reassessment used (e.g. physical exam, bedside cardiovascular ultrasound, passive leg raising, CVP, ScV02 assessment). The clinician can attest that volume and perfusion reassessment has occurred, even without reference to the method used. This will meet the measure's volume and perfusion reassessment requirement. A provider may also opt to state their chosen method, but this is not required.
7. All of the above numerator components (as applicable) must be in compliance, otherwise the case is calculated as a 'failed' sepsis case.

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.

The approach outlined below can also be found in the Measure Information Form attached to the submission. Additionally, the approach outlined below aligns with other measures which are part of the Inpatient Quality Reporting Program.

Sampling:

Hospitals have the option to sample from their population, or submit their entire population. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter/month for the measure cannot sample.

Population and Sampling:

Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter/month cannot sample. Hospitals that have five or fewer sepsis discharges for the entire measure set (both Medicare and non-Medicare combined) in a quarter are not required, but are encouraged to submit sepsis patient level data to the CMS Clinical Warehouse.

Quarterly Sampling:
Hospitals selecting sample cases for the sepsis measure must ensure that the population and quarterly sample size meets the following conditions:

Average Quarterly Initial Patient Population Size "N"	Minimum Required Sample Size
>=301	60
151 - 300	20% of Initial Patient Population size
30 - 150	30
6 - 29	No sampling; 100% Initial Patient Population
0 - 5	Submission of patient level data is encouraged but not required. If submission occurs, 1 – 5 cases of the Initial Patient Population may be submitted

Monthly Sampling:
Hospitals selecting sample cases for the sepsis measure must ensure that the population and monthly sample size meets the following conditions:

Average Monthly Initial Patient Population Size "N"	Minimum Required Sample Size
>=101	20
51 - 100	20% of Initial Patient Population size
10 - 50	10
< 10	No sampling; 100% Initial Patient Population

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.
N/A. The measure does not use survey or patient reported data.

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

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 data collection software are available for sale from vendors. Alternatively, facilities can download the free CMS Abstraction & Reporting Tool (CART). Paper tools for manual abstraction, which are posted on www.QualityNet.org, are also available for the CART tool. These tools are posted on www.QualityNet.org.

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)*
Hospital
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.)*
N/A. This measure has one set of specifications and does not have separate calculations of individual performance measures.

2. Validity – See attached Measure Testing Submission Form
NQF_Testing_Attachment_v7_0_fb_-004-.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.)
Yes

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

PREVIOUS TESTING

NATIONAL QUALITY FORUM

Measure missing data in Composite 2.0 from Composite 1.0

NQF #: 0500 **NQF Project:** Infectious Disease Project

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability*.)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
The Surviving Sepsis Campaign database provides unequivocal results that compare reliability across 200 organizations using the "all or nothing" composite measure specified in 2a1.

- Surviving Sepsis Campaign Database

Between January 2005 and March 2008, 15,775 subjects at 252 qualifying sites (individual hospitals) were entered into the Surviving Sepsis Campaign (SSC) database. Excluding hospitals that contributed fewer than 20 subjects, the final sample consisted of 15,022 patients at 165 hospitals (median, 57; range, 20–471 subjects per hospital). Data from up to eight quarters were analyzed from each site. Hospitals contributed data for a mean duration of 15.6 months (median, 14 months). Data from 15,022 subjects at 165 sites were analyzed to determine the compliance with bundle targets and association with hospital mortality.

Sites were instructed to set up screening procedures to identify patients with severe sepsis and septic shock based on previously established criteria as provided in a manual that included specific specifications consistent with those in section 2a1 of this submission.

Sites were provided a sample screening tool in the Campaign manual and on the Web site. Participating sites were asked to screen for patients in the emergency department, the clinical wards, and the ICU. To be enrolled, a subject had to have a suspected site of infection, 2 systemic inflammatory response syndrome criteria, and 1 organ dysfunction criterion. Clinical and demographic characteristics and time of presentation with severe sepsis criteria were collected for analysis of time-based measures.

Data were entered into the SSC database locally at individual hospitals into pre-established, structured data fields documenting performance and the time of specific actions and findings. Hospitals could not modify these fields. Hospitals were instructed to exclude only those patients with comfort measures status. These instructions were consistent with the exclusions identified in 2a1.8 of this submission and conveyed to sites in the manual that accompanied download of the database. Variation in structured data fields was not possible with the database and full completion was required to submit any data to the master database.

Data stripped of private health information were submitted every 30 days to the secure master SSC server at the Society of Critical Care Medicine via file transfer protocol or as comma delimited text files attached to e-mail submitted to the Campaign's server.

Additional data was obtained after the initial data collection period described above. The Surviving Sepsis Campaign database now contains data submitted from January 2005 through July 2012. Analyses constrained by the same criteria as above now are possible with a total of 28,150 patients with severe sepsis and septic shock at 218 international sites.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
The analytic methods for reliability testing of the data is described below:

- Surviving Sepsis Campaign Database

For purposes of reliability testing, the SSC data was analyzed as described in the RAND Corporations "The Reliability of Provider Profiling: A Tutorial" by John L. Adams (see appendix). This methodology is specifically endorsed by the NQF to analyze the reliability of performance for proposed metrics.

The analysis is sometimes referred to as a signal-to-noise analysis. As a measure of reliability, the analysis is a key determinant of suitability because it permits an assessment of how well one can confidently distinguish the performance of one entity from another. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance.

The SSC biostatician, Gary Phillips, MAS (Ohio State University Center for Biostatistics), used the Rand model to generate the reliability of the "all or nothing" composite metric specified in section 2a1 above by hospital site. The strategy involves fitting a beta-binomial model for each indicator. From each model two parameters are generated (alpha and beta) that define the beta-binomial distribution. From these parameters Mr. Philips then produced the between hospital variance. Next the within hospital variance is generated based on a proportion affirmative answers for each quality indicator (the binomial distribution). Analyzing the between hospital variance and the within hospital variance generates the reliability for each hospital site.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
The testing results for the reliability of the indicator are described below:

- Surviving Sepsis Campaign Database

The reliability of the measure using the SSC database (on a scale of 0 to 1) was examined by each site for the resuscitation bundle as described in 2a2.2 above. The mean reliability (and its associated standard deviation) for each of the deciles of reliability for the composite measure is as written in text format below.

A limitation of this online submission form is that tables cannot be easily inserted, therefore please see the text Table 1 "Reliability deciles with SD from beta-binomial model by site" just below. For specific reliabilities for each of 210 hospitals/sites with the associated number of contributed charts per hospital see Table 2 "Reliability estimated from beta-binomial model by site ID" included at the bottom of this field.

Table 1 "Reliability deciles with SD from beta-binomial model by site":

1st Decile: N = 22, Mean reliability = 0.751, SD of reliability = 0.0539.
2nd Decile: N = 22, Mean reliability = 0.850, SD of reliability = 0.0182.
3rd Decile: N = 22, Mean reliability = 0.889, SD of reliability = 0.0075.
4th Decile: N = 22, Mean reliability = 0.915, SD of reliability = 0.0077.
5th Decile: N = 21, Mean reliability = 0.929, SD of reliability = 0.0025.
6th Decile: N = 22, Mean reliability = 0.939, SD of reliability = 0.0049.
7th Decile: N = 22, Mean reliability = 0.956, SD of reliability = 0.0052.
8th Decile: N = 22, Mean reliability = 0.973, SD of reliability = 0.0041.
9th Decile: N = 22, Mean reliability = 0.988, SD of reliability = 0.0040.
10th Decile: N = 21, Mean reliability = 0.999, SD of reliability = 0.0013.

Total: N = 218, Mean reliability = 0.919, SD of reliability 0.0732.

Note also that although for purposes of this submission only the composite measure is being considered for endorsement, the specific reliabilities of the underlying components is known and were calculated using the same methodology.

Results are summarized below as percentages:

N = 28150

Serum lactate.....Mean reliability = 94.96.....SD = 4.57
Culture before antibiotics.....Mean reliability = 90.19.....SD = 8.03
Timely antibiotics.....Mean reliability = 87.66.....SD = 8.82

Fluids & Vasopressors.....Mean reliability = 94.12.....SD = 5.32
CVP Assessment.....Mean reliability = 90.81.....SD = 8.15
ScVO2 Assessment.....Mean reliability = 89.84.....SD = 89.84

Overall Bundle.....Mean reliability = 91.86.....SD = 7.30

Table 2 “Reliability estimated from beta-binomial model by site ID":

ID.....N.....Composite measure reliability

1.....63.....0.937
2.....38.....0.930
3.....36.....0.864
4.....34.....0.954
5.....34.....0.880
6.....73.....0.881
7.....35.....1.000
8.....50.....0.922
9.....21.....1.000
10.....49.....0.919
11.....62.....0.985
12.....66.....0.899
13.....97.....0.924
14.....67.....0.976
15.....46.....0.974
16.....22.....1.000
17.....43.....0.879
18.....40.....0.885
19.....29.....0.887
20.....63.....0.960
21.....37.....0.926
22.....47.....0.880
23.....52.....0.961
24.....37.....0.777
25.....27.....0.924
26.....22.....0.898
27.....29.....0.887
28.....27.....1.000
29.....82.....0.961
30.....35.....0.832
31.....28.....0.934
32.....50.....0.922
33.....29.....1.000
34.....92.....0.938
35.....31.....1.000
36.....34.....0.880
37.....94.....1.000
38.....217.....0.991
39.....96.....0.947
40.....80.....0.916
41.....20.....1.000
42.....28.....0.880

43.....20.....1.000
44.....40.....0.845
45.....25.....0.855
46.....73.....1.000
47.....44.....0.923
48.....38.....0.901
49.....306.....0.969
50.....28.....0.691
51.....62.....0.897
52.....230.....0.968
53.....285.....0.953
54.....214.....0.937
55.....20.....0.732
56.....58.....0.916
57.....183.....0.932
58.....126.....0.911
59.....30.....0.701
60.....62.....0.888
61.....684.....0.983
62.....40.....0.813
63.....70.....0.830
64.....52.....0.809
65.....310.....0.973
66.....20.....0.604
67.....59.....0.808
68.....39.....0.880
69.....75.....1.000
70.....27.....0.731
71.....169.....0.938
72.....304.....0.974
73.....100.....0.928
74.....204.....0.938
75.....159.....0.932
76.....101.....0.948
77.....97.....0.924
78.....60.....0.891
79.....50.....0.939
80.....85.....0.855
81.....117.....0.928
82.....110.....0.939
83.....87.....0.932
84.....167.....0.959
85.....36.....0.769
86.....22.....0.822
87.....57.....0.882
88.....110.....0.895
89.....44.....1.000
90.....34.....0.954
91.....236.....0.978
92.....108.....0.961
93.....47.....0.932
94.....93.....0.981
95.....165.....0.960
96.....27.....0.929

97.....27.....0.872
98.....39.....1.000
99.....23.....0.906
100.....33.....1.000
101.....140.....0.956
102.....264.....0.992
103.....114.....0.979
104.....136.....0.980
105.....312.....0.994
106.....20.....0.794
107.....203.....0.968
108.....27.....0.757
109.....62.....0.947
110.....40.....1.000
111.....67.....0.826
112.....351.....0.966
113.....37.....0.846
114.....162.....0.919
115.....155.....0.955
116.....148.....0.990
117.....58.....1.000
118.....528.....0.974
119.....105.....0.928
120.....20.....0.794
121.....58.....0.916
122.....240.....0.957
123.....87.....0.947
124.....177.....0.988
125.....414.....0.997
126.....159.....0.989
127.....51.....0.979
128.....621.....0.987
129.....264.....0.980
130.....35.....1.000
131.....207.....0.972
132.....80.....0.932
133.....200.....0.992
134.....126.....0.929
135.....148.....0.933
136.....187.....0.930
137.....104.....0.898
138.....26.....0.816
139.....35.....0.956
140.....36.....0.864
141.....87.....0.932
142.....68.....0.878
143.....31.....0.899
144.....168.....0.933
145.....22.....0.822
146.....101.....0.944
147.....52.....0.863
148.....166.....0.928
149.....107.....0.932
150.....74.....0.945

151.....551.....0.982
152.....80.....0.916
153.....75.....0.901
154.....89.....0.986
155.....23.....0.779
156.....153.....0.928
157.....21.....0.749
158.....23.....0.701
159.....25.....0.919
160.....421.....0.981
161.....454.....0.970
162.....167.....0.994
163.....1131.....0.998
164.....106.....0.884
165.....142.....0.911
166.....251.....0.948
167.....222.....0.942
168.....70.....0.853
169.....291.....0.953
170.....152.....0.934
171.....66.....0.892
172.....61.....0.841
173.....312.....0.997
174.....119.....0.988
175.....151.....0.969
176.....124.....0.946
177.....127.....0.918
178.....112.....0.925
179.....313.....0.992
180.....321.....0.986
181.....23.....0.834
182.....142.....0.934
183.....204.....0.961
184.....75.....0.857
185.....40.....0.748
186.....157.....0.933
187.....243.....0.948
188.....383.....0.968
189.....394.....0.985
190.....506.....0.985
191.....1023.....0.991
192.....100.....0.876
193.....34.....0.914
194.....70.....0.989
195.....473.....0.972
196.....37.....0.896
197.....28.....0.662
198.....41.....0.794
199.....290.....0.973
200.....227.....0.945
201.....93.....0.900
202.....28.....0.929
203.....26.....0.924
204.....231.....0.996

205.....150.....0.954
206.....93.....0.959
207.....150.....0.913
208.....44.....0.851
209.....251.....0.973
210.....31.....0.755

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* **are consistent with the evidence cited in support of the measure focus (***criterion 1c)* **and identify any differences from the evidence:**
The measure is consistent with regard to the evidence cited in criterion 1c with respect to the measure focus, the target population and the exclusions.

Each will be detailed below:

- Measure Focus:

The "all or nothing" composite measure specifications directly pertain to the measure focus (the early management of patients with severe sepsis and septic shock) inasmuch as the measure is constrained to observations in first 6 hours of care.

The specifications are identical to the evidence cited in 1c. Each cited piece of literature was evaluated using a binary (affirmative or negative) response as regards the provision of specific components of care delivery in determining the composite "all or nothing" calculation. The health outcomes assessed in each study involved no intermediate clinical outcomes, but rather focused on mortality in each instance. The reliability data provided from the SSC was derived under these conditions and, in fact, makes up the largest body of the clinical evidence for the measure focus.

- Target Population:

The specified target population of the measure is entirely consistent across the evidence. All studies have looked at patients with severe sepsis and septic shock as defined by the the 2001 Society of Critical Care Medicine/European Society of Intensive Care Medicine/American College of Chest Physicians/American Thoracic Society/Surgical Infection Society consensus sepsis definition. The data provided as regards reliability here is not to the contrary in any material form.

- Exclusions:

Exclusions were uniform from the specifications cited in 2a1 and the evidence cited here as regards reliability. In particular, patients were excluded in the SSC analysis with advanced directives for comfort care and/or clinical conditions that precluded total measure completion. These exclusions were conveyed to hospitals in the manual that accompanied the SSC database.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**
Although there may often be a distinction between the data used to inform the evidence for the importance of the measure focus and the data that informs reliability and validity testing, the authors of this submission have access to the database that produced the foundational evidence for the importance of the measure and have used the same data to produce calculations related to the reliability and validity of the data. Therefore, the distinction is not applicable in this case.

The data sample then is identical to the data identified in 2a2.1 and repeated here:

- Surviving Sepsis Campaign Database

Between January 2005 and March 2008, 15,775 subjects at 252 qualifying sites (individual hospitals) were entered into the Surviving Sepsis Campaign (SSC) database. Excluding hospitals that contributed fewer than 20 subjects, the final sample consisted of 15,022

patients at 165 hospitals (median, 57; range, 20–471 subjects per hospital). Data from up to eight quarters were analyzed from each site. Hospitals contributed data for a mean duration of 15.6 months (median, 14 months). Data from 15,022 subjects at 165 sites were analyzed to determine the compliance with bundle targets and association with hospital mortality.

Sites were instructed to set up screening procedures to identify patients with severe sepsis and septic shock based on previously established criteria as provided in a manual that included specific specifications consistent with those in section 2a1 of this submission.

Sites were provided a sample screening tool in the Campaign manual and on the Web site. Participating sites were asked to screen for patients in the emergency department, the clinical wards, and the ICU. To be enrolled, a subject had to have a suspected site of infection, 2 systemic inflammatory response syndrome criteria, and 1 organ dysfunction criterion. Clinical and demographic characteristics and time of presentation with severe sepsis criteria were collected for analysis of time-based measures.

Data were entered into the SSC database locally at individual hospitals into pre-established, structured data fields documenting performance and the time of specific actions and findings. Hospitals could not modify these fields. Hospitals were instructed to exclude only those patients with comfort measures status. These instructions were consistent with the exclusions identified in 2a1.8 of this submission and conveyed to sites in the manual that accompanied download of the database. Variation in structured data fields was not possible with the database and full completion was required to submit any data to the master database.

Data stripped of private health information were submitted every 30 days to the secure master SSC server at the Society of Critical Care Medicine via file transfer protocol or as comma delimited text files attached to e-mail submitted to the Campaign's server.

Additional data was obtained after the initial data collection period described above. The Surviving Sepsis Campaign database now contains data submitted from January 2005 through July 2012. Analyses constrained by the same criteria as above now are possible with a total of 28,150 patients with severe sepsis and septic shock at 218 international sites.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:
Validity testing of the performance measure presented here is based on the measure as specified and submitted for endorsement. The SSC database was used to generate this analysis. The performance scores reported were generated for the hospitals using the specifications submitted for endorsement.

As noted in the Measure Testing Task Force report, validity testing (as with reliability testing) can be conducted at the level of the critical data elements or the performance measure score. The testing presented here is testing at the level of the performance measure score based on conceptual understanding of the measure and also the data that are available.

• As such, a reasonable hypothesis for validity testing of the performance measure score would be that those with higher scores on the composite performance measure should have a lower score on a risk-adjusted mortality measure.

To test this hypothesis, the SSC biostatician, Gary Phillips, MAS (Ohio State University Center for Biostatistics) built a random effects logistic regression where patient level observations were nested within a particular site (hospital). The regression model is adjusted for the variables shown below.

The model included the following variables (Organ Failure abbreviated OF):

Sepsis origin
.....ED (referent)
.....Ward
.....ICU
Geographic region
.....Europe
.....North America (referent)
.....South America
Cardiovascular OF
.....Lactate > 4 mmol/L
.....Cardiovascular OF with Lactate > 4 mmol/L

.....No hypotension (referent)
.....Hypotension with MAP < 65 mm Hg
.....Hypotension with MAP = 65 mm Hg
Received =20 ml/kg of crystalloid or equivalent
Received vasopressors
Pneumonia
UTI
Abdominal
Meningitis
Catheter
Device
Other infection
Renal organ failure
Hepatic organ failure
Hematologic organ failure
No mechanical ventilation and no pulmonary OF (referent)
No mechanical ventilation and pulmonary OF
Mechanical ventilation with plateau pressure < 30 cm H2O and no pulmonary OF
Mechanical ventilation with plateau pressure < 30 cm H2O and pulmonary OF
Mechanical ventilation with plateau pressure = 30 cm H2O independent of pulmonary OF
Hyperthermia (> 38.3° C) or (101.0° F)
Hypothermia (< 36° C) or (96.8° F)
Chills with rigor
Tachypnea (BPM > 20)
Leukopenia (WBC count < 4,000/µL)
Hyperglycemia (plasma glucose > 120 mg/dL)
Acutely alter mental status

• Separately, there is substantial face validity of the measure. The Measure Testing TF report indicates that NQF does allow for face validity of the performance measure score if it is systematically assessed. Here, the measure submitted for evaluation does not substantially differ from the components of the composite measure systematically assessed in the SSC peer reviewed publication "Levy MM, Dellinger RP, Townsend SR, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. Critical Care Medicine 201 0; 38:367-74." In fact, the analysis performed for this submission is the identical data set to the SSC data set. This publication demonstrated declining mortality associated with increased compliance. The specific results are provided in 2b2.3 for review.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)***:**
• In the random effects logistic regression model created specific for this submission by Mr. Phillips described in 2b2.2, the odds of hospital mortality are reduced 10% (odds ratio = 0.90, 95% CI: 0.83 to 0.97, p-value = 0.008) for patients that are compliant with the measure as described in sections 2a1.1 and 2a1.4 of this submission.

• The face validity of the measure specifications is bolstered by the publication "Levy MM, Dellinger RP, Townsend SR, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. Critical Care Medicine 201 0; 38:367-74," which systematically assessed a composite measure not materially different from the performance measure here. This publication demonstrated declining mortality associated with increased compliance. The specific results are recited here:

"Outcome measures included hospital mortality, hospital length of stay, and ICU length of stay. Ten performance measures were established, based on the individual elements of the resuscitation bundle and the management bundle. The analysis set was constructed from the subjects entered into the SSC database from its launch in January 2005 through March 2008. The a priori data analysis plan limited inclusion to sites with at least 20 subjects and at least 3 months of subject enrollment. Analysis presented here was limited to the

first 2 yrs of subjects at each site. Sites were characterized by: hospital size (250, 250–500, >500 beds); teaching status; ICU type (medical, medical/surgical, other); and geographic region (Europe, North America, South America). Subjects were characterized by baseline severe sepsis information: location of enrollment (emergency department, ICU, ward); site of infection (pulmonary, urinary tract, abdominal, central nervous system, skin, bone, wound, catheter, cardiac, device, other); acute organ dysfunction (cardiovascular, pulmonary, renal, hepatic, hematologic). Data were organized by quarter through 2 yrs, with the first 3 months that a site entered subjects into the database defined as the first quarter, regardless of when those months occurred from January 2005 through March 2008. The effects of predictor variables on hospital mortality we expressed using odds ratios (ORs), including 95% confidence intervals (CIs) for risk-adjusted results. Logistic regression model fit was assessed using the Hosmer Lemeshow C statistic, the chi-square dispersion, the proportion of log-likelihood accounted for by the model, and an examination of model residuals. We constructed the databases in Access and Fox-Pro (Microsoft Corp, Redmond, WA) and conducted analyses in DataDesk (Data Description, Ithaca, NY) and SAS (SAS Institute, Cary, NC).

SURVIVING SEPSIS CAMPAIGN - CHANGES IN ACHIEVEMENTS OF BUNDLE TARGETS OVER TIME:

Compliance rates for achieving all bundle targets over time—both the overall bundles and the individual elements within both bundles—increased over time, although both basal achievement rates and the magnitude of improvement varied considerably across targets. Compliance with the initial bundle targets increased linearly from 10.9% of subjects in the first site quarter to 31.3% by the end of 2 yrs in the campaign, achieving statistical significance by the second quarter (10.9% vs. 14.9%, p .0001). The ability to achieve the entire bundle targets started higher, at 18.4% in the first quarter, and increased to 25.5% by the end of 2 yrs, but did not achieve statistical significance until the fourth quarter (18.4% vs. 21.5%, p < .008).

SURVIVING SEPSIS CAMPAIGN - CHANGES IN HOSPITAL MORTALITY:

Unadjusted hospital mortality decreased from 37.0% in the first quarter in the Campaign to 30.8% by 2 yrs (p < .001). On average, unadjusted mortality decreased by 0.91% (95% CI, 0.42–1.40) for each quarter in the Campaign. The results of the multivariable model examining the effect of time in the Campaign on hospital mortality, fit well (Hosmer and Lemeshow C statistic of 18.1 with 18 df, p <.34, accounted for 36.6% of variation in the data, with a chi-square dispersion of 1.04). In both the unadjusted and adjusted models, the chance of death decreased the longer a site was in the Campaign, resulting in an adjusted absolute drop of 0.8% per quarter and 5.4% over the first 2 yrs (95% CI, 2.5–8.4).

SURVIVING SEPSIS CAMPAIGN - RELATIONSHIP BETWEEN BUNDLE ELEMENTS AND IN HOSPITAL MORTALITY:

After adjustment for baseline characteristics, administration of broad-spectrum antibiotics (OR, 0.86; 95%, CI 0.79– 0.93; p .0001), obtaining blood cultures before their initiation (OR, 0.76; 95% CI, 0.70– 0.83; p < .0001) were all associated with lower hospital mortality. To control for entry of less severely ill patients in the database over time as the reason for decreasing mortality, severity was assessed based on variables linked to patient mortality that were available in the database. When mortality was adjusted accordingly, while the magnitude of the effect was slightly reduced, it remained statistically significant. The results of this study demonstrate that the use of a multifaceted performance improvement initiative was successful in changing sepsis treatment behavior as demonstrated by a significant increase in compliance with sepsis performance measures. This compliance was associated with a significant reduction in hospital mortality in patients with severe sepsis and septic shock. These results are consistent with an earlier report from Ferrer et al in Spain. The findings of this study show that the improvement in achievement of bundle targets and association with improved outcome is sustained over time and is demonstrated across a wide number of countries and settings."

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

If the component measures are combined at the patient level, complete 2b
2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
The validity analysis provided above provides discriminatory power at the level of the performance measure. Please note that the exclusions are highly intuitive and reasonable. Thus, imagining a world for testing purposes, where patients who did not consent for lines received them or who wished to be made comfort measures were treated aggressively is wholly unlikely. Exclusions referenced in section 2a1.8 therefore were not independently analyzed for their effect on the performance measure score given the impropriety of such testing. Such a study most likely could not be conducted due to appropriate IRB constraints.

Excluded patients are not included in data collection at the level of the data collection tool during the time of chart review (either in the sepsis campaign database or the paper equivalent tools, the "ICMT"). See 2a1.25 for specific algorithm and logic.

2b3.2 **Analytic Method** (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:
Not applicable, see 2b3.1.

2b3.3 **Results** *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
Not applicable, see 2b3.1.

If the component measures are combined at the patient level and include outcomes, complete 2e
2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 **Data/Sample** *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This process measure composite is not risk adjusted.

2b4.2 **Analytic Method (***Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:
This process measure composite is not risk adjusted.

2b4.3 **Testing Results** *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:
Not applicable.

2b4.4 **If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:** As a process measure, no risk adjustment is neccessary.

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 **Data/Sample** *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
The data sample used in this calculation of meaningful differences in performance (performed specifcially for this submission)was the enlarged SSC database (28,150 patients) identified in 2a2.1 and repeated here:

- Surviving Sepsis Campaign Database

Between January 2005 and March 2008, 15,775 subjects at 252 qualifying sites (individual hospitals) were entered into the Surviving Sepsis Campaign (SSC) database. Excluding hospitals that contributed fewer than 20 subjects, the final sample consisted of 15,022 patients at 165 hospitals (median, 57; range, 20–471 subjects per hospital). Data from up to eight quarters were analyzed from each site. Hospitals contributed data for a mean duration of 15.6 months (median, 14 months). Data from 15,022 subjects at 165 sites were analyzed to determine the compliance with bundle targets and association with hospital mortality.

Sites were instructed to set up screening procedures to identify patients with severe sepsis and septic shock based on previously established criteria as provided in a manual that included specific specifications consistent with those in section 2a1 of this submission.

Sites were provided a sample screening tool in the Campaign manual and on the Web site. Participating sites were asked to screen for patients in the emergency department, the clinical wards, and the ICU. To be enrolled, a subject had to have a suspected site of infection, 2 systemic inflammatory response syndrome criteria, and 1 organ dysfunction criterion. Clinical and demographic characteristics and time of presentation with severe sepsis criteria were collected for analysis of time-based measures.

Data were entered into the SSC database locally at individual hospitals into pre-established, structured data fields documenting performance and the time of specific actions and findings. Hospitals could not modify these fields. Hospitals were instructed to exclude only those patients with comfort measures status. These instructions were consistent with the exclusions identified in 2a1.8 of this submission and conveyed to sites in the manual that accompanied download of the database. Variation in structured data fields was not possible with the database and full completion was required to submit any data to the master database.

Data stripped of private health information were submitted every 30 days to the secure master SSC server at the Society of Critical Care Medicine via file transfer protocol or as comma delimited text files attached to e-mail submitted to the Campaign's server.

Additional data was obtained after the initial data collection period described above. The Surviving Sepsis Campaign database now contains data submitted from January 2005 through July 2012. Analyses constrained by the same criteria as above now are possible with a total of 28,150 patients with severe sepsis and septic shock at 218 international sites.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)***:**
For purposes of this analysis, the enlarged SSC database was used with 218 sites and 28,150 patients. The data set therefore covers 16 quarters of participation in the campaign, or 4 years worth of data total. Sites were permitted to join the campaign during the duration of the 4 years.

[NB, although the large majority of hospital sites joined early and some dropped out over time, few sites joined late. Given the discrepancy therefore between calendar time and time of participation in the campaign, sites were aligned by "site quarter" meaning that the first quarter of participation was the same for all sites regardless of the calendar month they joined the campaign. This pattern was maintained for all sites through 16 quarters. Of critical note, in the campaign to adjust for the confounding variable that mortality may be decreasing over time the mortality rate was determined at site quarter 1 of all participants and there was no statistical correlation with a decrease in mortality over time at the outset of participation. Thus, an underlying secular trend to lower mortality did not confound the data regardless of the variability in time when joining the campaign. The adjustment method and results of this important analysis are available in "Levy MM, Dellinger RP, Townsend SR, et al. The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. Critical Care Medicine 201 0; 38:367-74."]

Using the above data set, for this analysis, sites were aligned by number of site quarters of participation 1st quarter, 8th quarter (midpoint) and 16th quarter (endpoint). The total number of sites in the campaign at each time point was determined. The mean, SD, minimum, maximum, p25, p50, p75 of sites with a "yes" value for the composite indicator (i.e., a value counted in the numerator) is reported for each time referent in Table 1 below. Table 2 below reports the number of sites with decreasing performance over time compared with the number of sites with increasing performance over time. Please note the sharp attrition in sites reporting by site quarter 18, accounting for some of the declining performance detected. Efforts to promote participation in data collection had substantially fallen off by this time.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
Table 1. Descriptive summary of proportion where the composite is ´yes´ by site quarters of participation:

Quarters.of....Number....Mean....SD.....Min.....P25.....P50.....P75.....Max..
participation..of sites

1.....218.....0.101...0.180.....0.000...0.000...0.000...0.125...1.000
8.....88.....0.205...0.197.....0.000...0.000...0.183...0.354...0.667
16.....8.....0.158...0.102.....0.000...0.096...0.136...0.255...0.295

Please note as regards Table 1:

- Site quarters are based on a quarter of participation based on when a site entered the SSC program
- Site quarters do not align with calendar quarters (sites may enter variably)
- There were 218 sites that started the program
- After 2 years (8 quarters) there were 88 sites still participating
- After 4 years (16 quarters) there were 8 sites still participating

Table 2. Descriptive summary of proportion where the composite is ´yes´ by site quarters of participation:

Delta.....No.....Description..Mean...SD.....Min...P25....P50....P75....Max..
.....sites.....
.....
Decreased.....21.....Proportion...0.108..0.176..0.000..0.000..0.000..0.118..0.595
.....Delta.....0.247..0.246..0.021..0.059..0.200..0.334..1.000
.....
Increased.....67.....Proportion...0.235..0.196..0.000..0.000..0.213..0.385..0.667
.....Delta.....0.175..0.164..0.000..0.000..0.158..0.326..0.600

Please note as regards Table 2:
- Delta is whether or not a site decreased or increased
- 21 sites (24%) decreased from the 1st quarter while 67 sites (76%) increased
- Description: Proportion is those that met the resuscitation bundle and delta is the size of either the increase or decrease.

Finally, a description of the 8 sites with observations during site quarter 16:

- All 8 improved from the 1st quarter to the 8th quarter the mean proportion was 0.153 and the mean change from 1st to 8th quarter was 0.136
- Then from the 8th to the 16th quarter 3 sites decreased a mean proportion of 0.131 and 5 sites increased with a mean proportion of 0.087

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Please see attachment under "Supplemental Information" below.

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:
Please see attachment under "Supplemental Information" below.

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:
Please see attachment under "Supplemental Information" below.

2c. Disparities in Care: H☐ M☐ L☐ I☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: The Henry Ford Hospital, SCCM, IDSA, an emergency physician, and quality improvement organizations encourage the results of this measure to be stratified by race, ethnicity, gender, and primary language, and have included these variables as recommended data elements to be collected.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:
The Henry Ford Hospital, SCCM, IDSA, an emergency physician, and quality improvement organizations advocate that performance measure data should, where possible, be stratified by race, ethnicity, and primary language to assess disparities and initiate subsequent quality improvement activities addressing identified disparities, consistent with recent national efforts to standardize the collection of race and ethnicity data. A 2008 NQF report endorsed 45 practices including stratification by the aforementioned variables.(1) A 2009 IOM report "recommends collection of the existing Office of Management and Budget (OMB) race and Hispanic ethnicity categories as well as more finegrained categories of ethnicity(referred to as granular ethnicity and based on one's ancestry) and language need (a rating of spoken English language proficiency of less than very well and one's preferred language for health-related encounters)."(2)

References:
(1)National Quality Forum Issue Brief (No.10). Closing the Disparities Gap in Healthcare Quality with Performance Measurement and Public Reporting. Washington, DC: NQF, August 2008.
(2)Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. March 2010. AHRQ Publication No. 10-0058-EF. Agency for Healthcare Research and Quality, Rockville, MD. Available at: http://www.ahrq.gov/research/iomracereport. Accessed May 25, 2010.

2i. Component Item/Measure Analysis to Justify Inclusion in Composite

2i.1. Data/Sample
Please see attachment under "Supplemental Information" below.

2i.2. Analytic Method
Please see attachment under "Supplemental Information" below.

2i.3. Result
Please see attachment under "Supplemental Information" below.

2j. Component Item/Measure Analysis of Contribution to Variability in Composite Score

2j.1. Data/Sample
Please see attachment under "Supplemental Information" below.

2j.2. Analytic Method
Please see attachment under "Supplemental Information" below.

2j.3. Result
Please see attachment under "Supplemental Information" below.

2k. Analysis to Support Differential Weighting of Component Score

2k.1. Data/Sample
Composite score is not weighted.

2k.2. Analytic Method
Composite score is not weighted.

2k.3. Result
Composite score is not weighted.

2k.4. Describe how the method scoring/aggregation achieves the stated purpose and represent the quality construct
Composite score is not weighted.

2k.5. Indicate if any alternative scoring/aggregation methods were tested and why not chosen
Composite score is not weighted.

2l. Analysis of Missing Component Scores

2l.1. Data/Sample
For missing component scores fail opportunity score.

2l.2. Analytic Method

For missing component scores fail opportunity score. 2I.3. **Result** For missing component scores fail opportunity score.
2.1-2.3 **Supplemental Testing Methodology Information:** Attachment NQF_0500_Tables_and_Forest_Plots.pdf
Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met? **(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐** **Provide rationale based on specific subcriteria:**
If the Committee votes No, STOP

THIS FORM HAS BEEN UPDATED SINCE DATE OF SUBMISSION

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

Measure Number (*if previously endorsed*)**:** 0500

Composite Measure Title: Early Management Bundle, Severe/Septic Shock

Date of Submission: 2/3/2017

Composite Construction:

☐Two or more individual performance measure scores combined into one score

☒ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)

Instructions: Please contact NQF staff before you begin.

- If a component measure is submitted as an individual performance measure, the non-composite measure testing form must also be completed and attached to the individual measure submission.
- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **Sections 1, 2a2, 2b2, 2b3, 2b5, 2b7, and 2c must be completed.**
- **For composites with outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b7) and composites (2c) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 7.0 of the Measure Testing Attachment and the 2016 Measure Evaluation Criteria and Guidance.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2.the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(*if not conducted or results not adequate, justification must be submitted and accepted)*

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions

across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☒ **abstracted from paper record**	☒ **abstracted from paper record**
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☒ **abstracted from electronic health record**	☒ **abstracted from electronic health record**
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

Reliability Testing and Measure Score Validity Testing is based upon a sample of SEP-1 chart-abstracted data submitted to the CMS Clinical Data Warehouse (through the *QualityNet Secure Portal*) by hospitals participating in the CMS Hospital Inpatient Quality Reporting Program.

Data Element Validity Testing is based upon validated patient-level data submitted to the CMS Clinical Data Warehouse for the SEP-1 measure by hospitals participating in the CMS Hospital Inpatient Quality Reporting Program.

Exclusions Analysis is aggregated measure-level data submitted to the CMS Clinical Data Warehouse for the SEP-1 measure by hospitals participating in the CMS Hospital Inpatient Quality Reporting Program.

1.3. What are the dates of the data used in testing? 10/1/2015- 6/30/2016

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice

☒ **hospital/facility/agency**	☒ **hospital/facility/agency**
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which <u>measured entities</u> 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*)

Reliability Testing

For reliability testing, we used a random sample (determined using the sampling methodology described in Submission Form S.15) of SEP-1 chart-abstracted data submitted to CMS as part of the Hospital Inpatient Quality Reporting Program for October 2015 to June 2016 (Q4 2015 to Q2 2016). For SEP-1, the data included 3,134 to 3,193 hospitals, depending on the quarter. The sample included 302,281 cases in the denominator after removing exclusions and 119,048 numerator cases.

Validity Testing – Measure Score Validity

Since the last submission, the NQF-0500 measure was introduced as SEP-1, the first CMS sepsis measure with reporting beginning in October, 2015. The inpatient Quality Reporting (IQR) Program Performance requirements were met by 99% of the reporting hospitals. Using SEP-1 data from the Clinical Data Warehouse (CDW), CMS conducted a time trend analysis to assess the mortality rate across 159,289 patients who were eligible for the SEP-1 measure from October 2015 to June 2016.

Validity Testing – Data Element Validity

During Q4 2015 (October 2015 – December 2015), a sample of acute care hospitals were randomly selected for validation from all the care hospitals across the United States who submitted cases to the CMS Clinical Data Warehouse data set for the SEP-1 measure for the CMS Hospital Inpatient Quality Reporting Program. For 4th quarter 2015, 281 hospitals were selected for validation.

1.6. How many and which <u>patients</u> 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*)

Reliability Testing

Not applicable. The reliability analysis was conducted at the facility level and did not use patient-level information.

Validity Testing – *Data Element Validity*

The test population selected for validation represents a random sample of 303 cases submitted for the 4th quarter of 2015. No patient level information was reported.

Validity Testing – *Measure Score Validity*

The measure score validity analysis included 159,289 patients who were eligible for the SEP-1 measure from October 2015 to June 2016.

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.

Reliability Testing:

Data Source (submitted the Clinical Data Warehouse): Random sample (determined using the sampling methodology described in Submission Form S.15 submitted by each of the up to 3,193 providers
Dates: October 2015 – June 2016

Number of Facilities: 3,134 -3,193, depending on the quarter
Denominator cases after exclusions: 302,281 across all three quarters
Numerator Cases: 119,048 across all three quarters
Level of Analysis: Facility

Validity Testing – *Data Element Validity*
Data Source: Clinical Data Abstraction Center (CDAC) and Random sample of data submitted to the Clinical Data Warehouse.
Dates: October 2015 – December 2015
Number of Facilities: 281
Sample (cases): 303
Level of Analysis: Patient

Validity Testing – *Measure Score Validity*
Data Source: (submitted the Clinical Data Warehouse): Random sample (determined using the sampling methodology described in Submission Form S.15 submitted by each of the up to 3,193 providers
Dates: October 2015 – June 2016
Number of Facilities: 3,134 -3,193, depending on the quarter
Number of eligible cases: 159,289
Level of Analysis: Patient

Exclusion Testing:
Data Source (submitted to Clinical Data Warehouse): Random sample (determined using the sampling methodology described in Submission Form S.15 submitted by each of the up to 3,193 providers
Dates: October 2015 – June 2016
Number of Facilities: 3,134 -3,193, depending on the quarter
Exclusions: 339,678
Level of Analysis: Measure-level, aggregated

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).

Patient level SDS factors were assessed and reported in Section 1.b.4, which provides an overview of disparities in care for patient sub-populations. While an analysis of SDS factors is important in understanding differences in care for patient sub-populations, this measure is a process measure that is neither risk-adjusted nor risk-stratified. We determined that risk adjustment and risk stratification were not appropriate based on the current evidence base and the measure construct.

2a2. RELIABILITY TESTING

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
Note*: Current guidance for composite measure evaluation states that reliability must be demonstrated for the composite performance measure score.*
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. Describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

We calculated the measure's reliablity score in accordance with the method discussed in The Relialbity of Provider Profiling: A Tutorial (2009). The approach calculates the measure's ability to distinguish true performce diffrences among

different facilities. Higher reliability scores indicate greater distinguishability among different facilities.
The median reliability score was calculated under two scenarios: 1) the score was first calculated by including all facilities even if they had only one reported case in the given data period; 2) the score was calculated with a 10-case minimum cut-off, CMS's minimum requirement for public reporting, where facilities with fewer cases were dropped from the analysis. After applying the cutoff, more than 86% of the facilities remained for the reliability calculation. The minimum requirement exists to mitigate the issue of sampling bias on the overall analysis. The results of the reliability testing are reported by quarter.
We used quarterly data covering the Oct 2015 – June 2016 data collection period. The reliability score is estimated using beta-binomial model, which is appropriate for testing the reliablity of pass/fail measures. The reliablity score for each facility is a function of the provider's performance rate, the faciliity's sample size, and the variance across facilities.
REFERENCE:
1) Adams JL. The reliablity of provider profiling: a tutorial. Santa Monica, CA: RAND Corporation. 2009. Retrieved from http://www.rand.org/pubs/technical_reports/TR653.

2a2.3. 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*)

Using all reported cases, the reliability score for the measure was 0.92 (CI 0.41 – 1.00, 2015 Q4), 0.93 (CI 0.47 – 1.00, 2016 Q1), and 0.93 (CI 0.42 – 1.00, 2016 Q2). The confidence interval gets narrower from 2015 Q4 to 2016 Q1, but remains consistent in 2016 Q2.

With the minimum threshold for cases reported set at 10 cases per quarter, we see similar reliability scores to all cases. However, the confidence intervals with the minimum threshold for cases become noticeably narrower, across all quarters, as compared to using all cases, signaling more confidence in the results of the reliability testing. Using the minimum threshold for cases, he confidence interval is 0.63 – 0.99 for 2015 Q4, 0.64 – 0.99 for 2016 Q1, and 0.65 – 0.99 for 2016 Q2.

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

Calculated using a beta-binomial model, an average reliability score of 0.92 across all reporting quarters is indicative of high measure reliability. Additionally, as we filter out facilities with cases fewer than 10, the reliability results stay consistent, but with a narrowing of the confidence interval. This change in the confidence interval with the implementation of a minimum threshold further bolsters the high measure reliability results.

2b2. VALIDITY TESTING

Note*: Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*

2b2.1. What level of validity testing was conducted?

☒ **Composite 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*)
- ☐ **Systematic assessment of content validity**

☒ **Validity testing for component measures** *(check all that apply)*

Note*: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*
- ☐ **Endorsed (or submitted) as individual performance measures**
- ☒ **Critical data elements** (*data element validity must address ALL critical data elements*)
- ☐ **Empirical validity testing of the component measure score(s)**

☐ **Systematic assessment of face validity of <u>component measure score(s)</u> 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)*

Validity Testing – *Measure Score Validity*:

Sepsis Mortality Analysis – Overall

The breakdown below includes all eligible cases in the mortality analysis population which includes the overall measure outcome (see Sepsis Mortality Analysis – Overall below). The Total Percentage calculation is derived from each of the number of cases in the population description divided by the respective outcome for the total number of cases in the overall measure. For the Total Percentage field, the total number of Excluded and Eligible Sepsis Cases are divided by the total number of the Initial Population. Then, the total number of Passed and Failed Sepsis Cases are each divided by the total number of Eligible Sepsis Cases.

The report below also includes multiple counts and percentages of deaths. The 'Total Deaths' field adds total deaths at discharge and up to 30 days after discharge together to show the cases that died on or sometime within 30 days of being discharged. Each of the death percentage fields are derived by taking the respective death counts and dividing by the number of cases for each of the different populations in the report.

We performed a Chi-Square Test of Association and Equal Proportions between the two categorical variables: Measure Outcome (Failed or Passed) and Mortality Result (Died or Survived) (see Chi-Square Test of Associate and Equal Proportions results below). For this test, a Chi-Square value was calculated along with its associated p-value and a Risk Ratio with a 95% Confidence Limit was derived.

Sepsis Rate Comparisons by Percentiles

The first approach was to use all cases that either passed or failed the SEP-1 Measure and calculate pass rates for the measure for each provider. The pass rate deciles was calculated based on the distribution of each provider pass rate. Using the calculated pass rate deciles, each provider was assigned into a percentile grouping based on their respective pass rates. The pass rates were re-calculated for each of the 10 percentiles along with the calculated mortality rates for each percentile. The number of providers and number of cases that fell into each of the percentiles are included in Table 1. and Chart 1. Sepsis Rate Comparisons by Percentiles (2015Q4 – 2016Q2).

The second approach was to use the same exact population of data as the first, but to use hard cut-offs for pass rates and assign providers to different pass rate buckets manually based on their respective pass rates and leaving the pass rate distributions out of the process. Again, the pass rates were re-calculated for each of the 10 hard cut-off pass rate buckets that were created along with the calculated mortality rates for each bucket. The number of providers and number of cases that fell into each of the hard cut-off pass rate buckets are included in Table 2. and Chart 2. Sepsis Rate Comparisons by Buckets (2015Q4 – 2016Q2).

The biggest difference between the two approaches is that by calculating pass rate deciles based on the distributions, it allows for more even counts of providers in each of the 10 percentiles. The hard cut-off pass rates will be much less balanced in the counts of providers due to leaving the pass rate distributions out of the process of assigning the cut-off pass rates for each of the buckets.

Two Proportion Z-Tests

A good rule of thumb when getting the P-value with the associated z-score, if the z-score is about +/- 1.64 then you have a p-value equal to about 0.05, which with a 95% significance level would mean you reject the null hypothesis and there is a statistically significant difference between the two tested proportions.

Validity Testing - *Critical data elements*:

Copies of medical records for each case selected are requested from the facilities selected for validation. An independent group of medical record abstractors from the CDAC under contract with CMS abstracts the same data elements for each case that the submitted hospital abstracted when originally submitting data to the CMS clinical data warehouse. The CDAC compares abstraction results and identifies cases and data elements for which there is a mismatch in abstraction. Using that information, we calculated a percent agreement between each of the facility-abstracted data elements and gold standard validated elements.

CDAC sampled 281 facilities for validation. Among these facilities, 303 cases were selected randomly to be validated.

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

Sepsis Mortality Analysis - Overall

Population Description	Cases	Total Percentage	Total Deaths	Total Deaths Percentage
Initial Population Number of Sepsis Cases	**325,809**	**--**	**81,587**	**25.0%**
Total Number of Excluded Sepsis Cases	*166,520*	*51.1%*	*38,624*	*23.2%*
Total Number of Eligible Sepsis Cases	**159,289**	**48.9%**	**42,963**	**27.0%**
Total Number of Passed Sepsis Cases	*64,051*	*40.2%*	*14,039*	*21.9%*
Total Number of Failed Sepsis Cases	*95,238*	*59.8%*	*28,924*	*30.4%*

Chi-Square Test of Associate and Equal Proportions

Population	Chi-Square	P-Value	Risk Ratio	Lower 95% Confidence Limit	Upper 95% Confidence Limit
Total Severe Sepsis Eligible Cases	1388.8163	<.0001	1.3856	1.3616	1.4101

Sepsis Rate Comparisons by Percentiles

Table 1. Sepsis Rate Comparisons by Percentiles (2015Q4 - 2016Q2)				
Pass Rate Percentiles	**Num Providers**	**Num Cases**	**Pass Rate**	**Mortality Rate**
10th	309	15,333	72.14%	24.27%
20th	300	16,119	58.78%	24.36%
30th	317	18,177	51.35%	25.60%
40th	311	19,139	44.94%	26.03%
50th	310	17,914	39.39%	26.45%
60th	306	18,598	34.38%	27.93%
70th	312	19,880	29.03%	28.47%
80th	307	15,927	23.49%	28.57%

90th	302	13,962	16.79%	29.95%
100th	320	4,240	6.44%	32.00%

Chart 1. Sepsis Rate Comparisons by Pass Rate Percentiles (2015Q4 - 2016Q2)

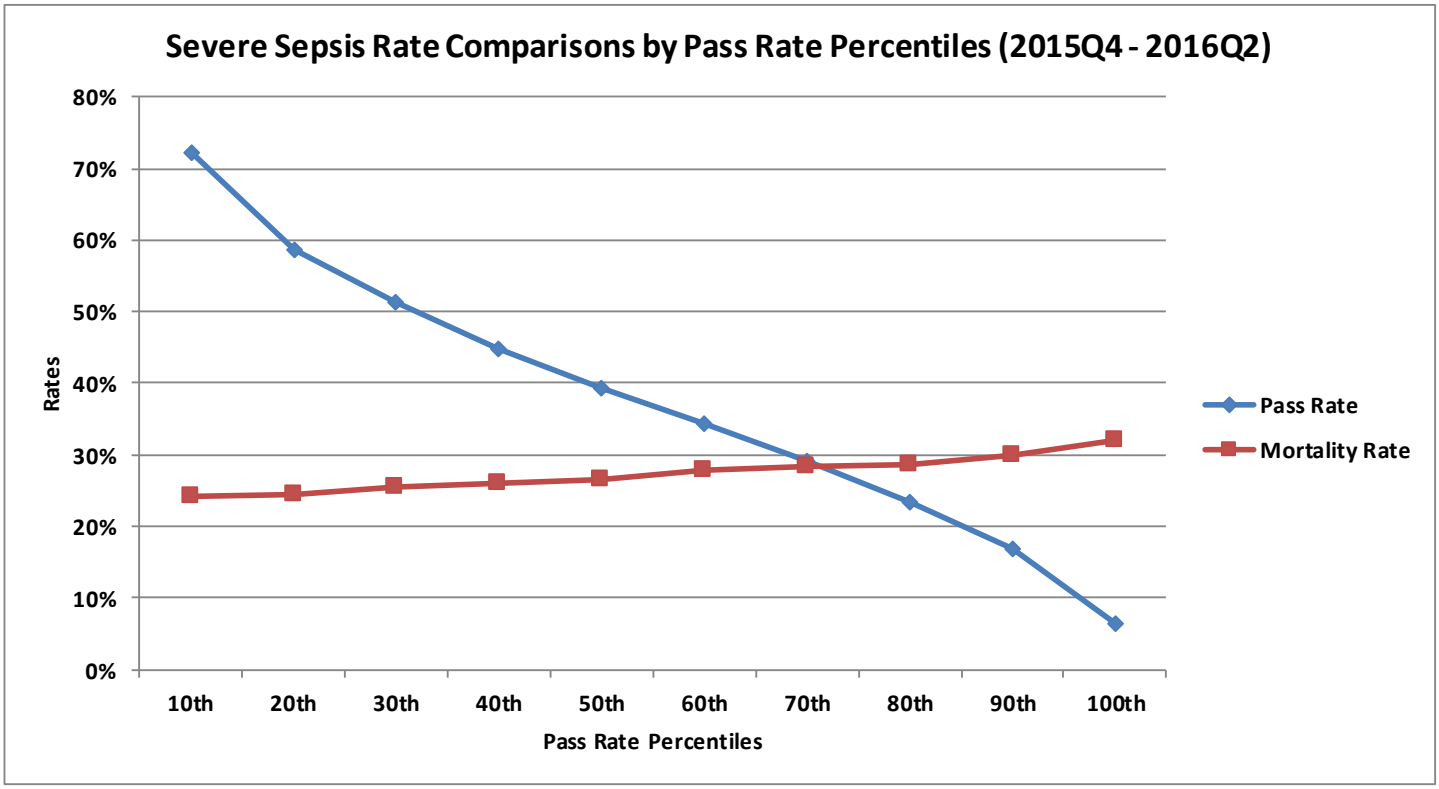


Chart 1. Sepsis Rate Comparisons by Pass Rate Percentiles (2015Q4 - 2016Q2), shows the overall Pass Rate compared to the overall Mortality Rate across each of the calculated provider pass rate percentiles.

Table 2. Sepsis Rate Comparisons by Buckets (2015Q4 - 2016Q2)

Pass Rate Buckets	Num Providers	Num Cases	Pass Rate	Mortality Rate
90.01% - 100.00%	36	325	96.62%	27.38%
80.01% - 90.00%	41	2,210	83.71%	23.03%
70.01% - 80.00%	103	5,228	73.60%	24.39%
60.01% - 70.00%	227	12,730	64.80%	23.97%
50.01% - 60.00%	376	22,845	54.72%	25.40%
40.01% - 50.00%	543	31,729	45.06%	25.62%
30.01% - 40.00%	642	36,813	35.03%	27.78%
20.01% - 30.00%	518	30,182	25.45%	28.53%
10.01% - 20.00%	315	13,858	16.15%	30.25%
0.00% - 10.00%	293	3,369	5.34%	32.03%

Chart 2. Sepsis Rate Comparisons by Pass Rate Buckets (2015Q4 - 2016Q2)

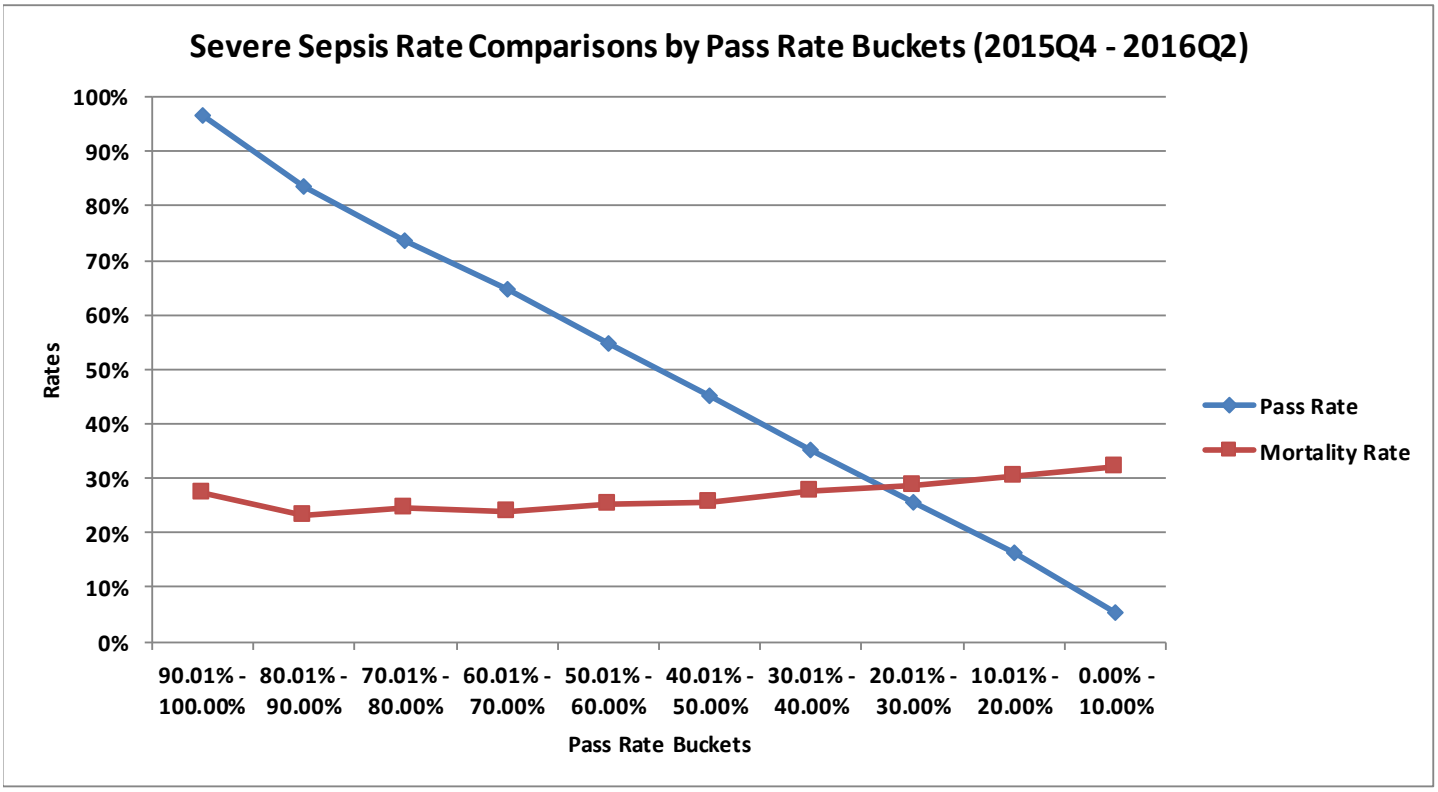


Chart 2. Sepsis Rate Comparisons by Pass Rate Buckets (2015Q4 - 2016Q2), shows the overall Pass Rate compared to the overall Mortality Rate across each of the manually assigned provider pass rates.

Additional analyses for each of the three quarters (2015Q4, 2016Q1, and 2016Q2) are included in the attached spreadsheet.

Two Proportion Z-Tests

NULL HYPOTHESIS: P1=P2 (MORTALITY Rate)					
P1	**P2**	**Pooled sample proportion**	**Standard error**	**Test statistic**	**P-value**
10th Pctl	20th Pctl	0.24316125	0.00483939	-0.18597396	0.4262
20th Pctl	30th Pctl	0.25017204	0.00468589	-2.64624408	0.0041
30th Pctl	40th Pctl	0.25820543	0.00453264	-0.94867429	0.1714
40th Pctl	50th Pctl	0.26233057	0.00457310	-0.91841330	0.1792
50th Pctl	60th Pctl	0.27203863	0.00465863	-3.17690227	0.0007
60th Pctl	70th Pctl	0.28208996	0.00459086	-1.17625098	0.1198
70th Pctl	80th Pctl	0.28514480	0.00480121	-0.20828090	0.4175
80th Pctl	90th Pctl	0.29214637	0.00527214	-2.61753333	0.0044
90th Pctl	100th Pctl	0.30427530	0.00806780	-2.54096680	0.0055

NULL HYPOTHESIS: P1=P2 (PASS Rate)					
P1	**P2**	**Pooled sample proportion**	**Standard error**	**Test statistic**	**P-value**
10th Pctl	20th Pctl	0.65293064	0.00537011	24.87844760	<0.0001
20th Pctl	30th Pctl	0.54842074	0.00538413	13.79981616	<0.0001
30th Pctl	40th Pctl	0.48062376	0.00517452	12.38761088	<0.0001

40th Pctl	50th Pctl	0.42256744	0.00513517	10.80783117	<0.0001
50th Pctl	60th Pctl	0.36838072	0.00504969	9.92140515	<0.0001
60th Pctl	70th Pctl	0.31615875	0.00474346	11.27868840	<0.0001
70th Pctl	80th Pctl	0.26565801	0.00469699	11.79479342	<0.0001
80th Pctl	90th Pctl	0.20360240	0.00466843	14.35170504	<0.0001
90th Pctl	100th Pctl	0.14379056	0.00615259	16.82218267	<0.0001

Data Element Validity Results:

Data Element	Number of Mismatches	Total Number of Questions Validated	Percent Agreement
Capillary Refill Examination Date	0	3	100.0%
Capillary Refill Examination Time	0	3	100.0%
Fluid Challenge Date	0	2	100.0%
Fluid Challenge Time	0	2	100.0%
Peripheral Pulse Evaluation Date	0	2	100.0%
Peripheral Pulse Evaluation Performed	0	3	100.0%
Peripheral Pulse Evaluation Time	0	2	100.0%
Skin Examination Date	0	2	100.0%
Skin Examination Performed	0	2	100.0%
Skin Examination Time	0	2	100.0%
Vasopressor Administration	0	6	100.0%
Transfer From Another Hospital or ASC	3	295	99.0%
Administrative Contraindication to Care	7	303	97.7%
Repeat Lactate Level Date	1	34	97.1%
Directive for Comfort Care, Severe Sepsis	3	38	92.1%
Bedside Cardiovascular Ultrasound Performed	1	9	88.9%
Central Venous Oxygen Measurement	1	9	88.9%
Central Venous Pressure Measurement	1	9	88.9%
Passive Leg Raise Exam Performed	1	9	88.9%
Vital Signs Review Date	1	8	87.5%
Vital Signs Review Time	1	8	87.5%
Septic Shock Presentation Date	5	38	86.8%
Discharge Disposition	20	151	86.8%
Cardiopulmonary Evaluation Date	1	7	85.7%
Cardiopulmonary Evaluation Time	1	7	85.7%
Severe Sepsis Present	46	268	82.8%
Crystalloid Fluid Administration	7	38	81.6%
Vasopressor Administration Date	1	5	80.0%
Vasopressor Administration Time	1	5	80.0%
Repeat Lactate Level Time	7	34	79.4%
Blood Culture Collection Date	25	120	79.2%
Initial Lactate Level Date	25	113	77.9%
Fluid Challenge Performed	2	9	77.8%

Broad Spectrum or Other Antibiotic Administration Date	32	142	77.5%
Directive for Comfort Care, Septic Shock	35	155	77.4%
Persistent Hypotension	5	21	76.2%
Broad Spectrum or Other Antibiotic Administration	36	149	75.8%
Initial Lactate Level Collection	36	149	75.8%
Blood Culture Collection	32	132	75.8%
Cardiopulmonary Evaluation Performed	2	8	75.0%
Initial Lactate Level Result	34	132	74.2%
Septic Shock Present	34	132	74.2%
Severe Sepsis Presentation Date	41	155	73.6%
Vital Signs Review Performed	3	11	72.7%
Repeat Lactate Level Collection	20	72	72.2%
Capillary Refill Examination Performed	2	7	71.4%
Blood Culture Collection Time	36	120	70.0%
Crystalloid Fluid Administration Date	8	26	69.2%
Broad Spectrum or Other Antibiotic Administration Time	47	142	66.9%
Broad Spectrum or Other Antibiotic Administration Selection	26	73	64.4%
Initial Lactate Level Time	41	113	63.7%
Discharge Time	5	13	61.5%
Septic Shock Presentation Time	17	38	55.3%
Crystalloid Fluid Administration Time	15	26	42.3%
Severe Sepsis Presentation Time	98	155	36.8%

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?*)

Measure Score Validity Results:

A Risk Ratio value higher than 1 with a significant p-value, would indicate that there is a higher risk of dying when a case fails the measure compared to when a case passes the measure. The Risk Ratio can be used as an actual ratio and it can be said that with 95% confidence, cases that fail the measure have 1.36 to 1.41 times the risk of dying compared to cases that pass the measure. It can also be used as a percentage and be said that with 95% confidence, cases that fail the measure have a 36% to 41% increase in risk of dying compared to cases that pass the measure.

Sepsis Rate Comparisons by Percentiles

Overall, the outcomes are very similar demonstrating a negative association between pass rates and the respective mortality rates.

Two Proportion Z-Tests

Four of the percentile comparisons have a statistically significant difference between mortality rates at a significance level of 0.05. There's also three additional percentile comparisons that are fairly close to a statistically significant difference between mortality rates at a significance level of 0.10. Where the difference is not clearly significant from one decile to the next, it does become significant adding in the next decile to the decile below.

Data Element Validity Results:

When using the percent agreement method to assess validity, results with better than 90% agreement are considered acceptable. Out of the 55 data elements tested for validity, 15 data elements (27.27%) have a percent agreement higher than 90%.

Accordingly, the top five data elements with the highest percentage of agreement are the capillary refill examination date and time, fluid challenge date and time and the peripheral pulse evaluation date. One explanation for the very high degree of agreement in these data elements would be the ease of abstracting the date and time elements, as well as the lower

frequency for conducting assessments like the capillary refill examination, fluid challenge, and peripheral pulse evaluation in the clinical setting.
Forty data elements (72.73 %) have a percentage agreement lower than 90%. In the process of interpreting this data, it is important to note the explanation for the lower percentage agreement is that the result of the validation efforts above represents data from Q4 2015. There have been numerous education and outreach efforts since Q4 2015 and updates to the measure with the intent of clarifying guidance and decreasing abstractor complexities, which would further improve on successive validation efforts.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

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*)

We tested measure exclusions to determine the prevalence of each exclusion at an aggregate level. The analysis tested measure exclusions during the October 1, 2015- June 30, 2016 data collection period. Measure exclusions include all cases meeting one or more criteria listed below:

Exclusion 1: Severe Sepsis is not present

Exclusion 2: Transfer from another hospital

Exclusion 3: Broad Spectrum Antibiotic was given more than 24 hours ago

Exclusion 4: Directive for comfort care, Severe Sepsis

Exclusion 5: Administrative Contraindication to care, Septic Shock

Exclusion 6: Administrative Contraindication to care, Severe Sepsis

Exclusion 7: Directive for comfort care, Septic Shock

Exclusion 8: Septic shock expired time less than 6 hours

Exclusion 9: Severe Sepsis expired time less than 3 hours

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*)

The data presented here is aggregated across 3 quarters from October 2015 to June 2016. Additionally, the Algorithm step noted in the first column aligns with the numbering in S.14 (Calculation Algorithm/Measure Logic) of the measure submission.

Algorithm	Data Element Name	Cases	Percent
	Total Number of Excluded SEP-1 Cases	339,678	100.00%
Step-04	Severe Sepsis Present is not present	245,737	72.34%
Step-03	Transfer from Another Hospital	62,502	18.40%
Step-22	Broad Spectrum Antibiotic Time > 24 hours	13,112	3.86%
Step-07	Directive for Comfort Care, Severe Sepsis	9,920	2.92%

Step-02	Administrative Contraindication to care, Severe Sepsis and Septic Shock	6,007	1.77%
Step-42	Directive for Comfort Care, Septic Shock	1,256	0.37%
Step-45	Shock Expired Time < 6 hours	620	0.18%
Step-11	Sepsis Expired Time < 3 hours	524	0.15%

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*)

The number of exclusions as shared above are not significant enough to unfairly distort measure performance results, and potentially negatively affect the reliability of the measure because the vast majority reflect cases that do not have severe sepsis and should not be analyzed. Additionally, excluding the 339,678 cases that are not relevant to determine performance on the SEP-1 measure reduces abstractor burden. Finally, there is strong clinical rationale and previous precedent across other measures (part of the Hospital Inpatient Quality Reporting program) to exclude cases with the above exclusions present.

_____

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

Note: *Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.*

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? (*check all that apply*)

☐ **Endorsed (or submitted) as individual performance measures**

☒ **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. This measure is not an outcome or a resource-use measure

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. This measure is not an outcome or a resource-use measure

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. This measure is not an outcome or a resource-use measure

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

N/A. This measure is not an outcome or a resource-use measure

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. This measure is not an outcome or a resource-use measure

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. This measure is not an outcome or a resource-use measure

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. This measure is not an outcome or a resource-use measure

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
N/A. This measure is not an outcome or a resource-use measure

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
N/A. This measure is not an outcome or a resource-use measure

2b4.9. Results of Risk Stratification Analysis:
N/A. This measure is not an outcome or a resource-use measure
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. This measure is not an outcome or a resource-use measure

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

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
Note: *Applies to the composite performance measure.*

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)*

Differences in performance rates for facilities meeting the requirements were tested. For the October 2015 – December 2015 data collection period, this included 3,134 facilities. For the January 2016 - March 2016 data collection period, this included 3,182 facilities. For the April 2016 - June 2016 data collection period, this included 3,193 facilities. Additional details of this analysis are provided in Section 2b5.2.

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*)

Below we include distribution information (facility-level distribution) on performance rates calculated for three quarters, between Q4 2015 and Q2 2016.
In each quarter, there is a wide range in hospital performance, indicating a continued gap in performance.

Q4 2015 Facility Level Analysis
3,134 hospitals submitted 96,516 eligible cases.
Min 0
10th percentile 5.0%
25th percentile 17.9%
Median 31.0%
75th percentile 45.8%
90th percentile 60.0%
Max 100.0%

Q1 2016 Facility Level Analysis
3,182 hospitals submitted 104,166 eligible cases.
Min 0
10th percentile 7.7%
25th percentile 21.6%
Median 36.1%
75th percentile 51.3%
90th percentile 66.7%
Max 100

Q2 2016 Facility Level Analysis
3,193 hospitals submitted 101,599 eligible cases.
Min 0
10th percentile 12.5%
25th percentile 25.8%
Median 41.7%
75th percentile 57.1%
90th percentile 71.4%
Max 100

Please also refer to the measure performance table in the measure submission form table appendix.

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 measure was able to detect facilities with better- and worse-than-average performance. The facility performance scores ranged from 0.0% to 100.0%, with a median rate ranging from 31.0% to 41.7% across the three quarters of available data. Fifty percent of facilities fell within the interquartile range of 17.9% to 45.8% in Q4 2015, 21.6% to 51.3% in Q1 2016, and 25.8% to 57.1% in Q2 2016. The SD performance rate ranged from 21.1% to 22.9%. Analysis of the October 2015-June 2016 performance data demonstrates the ability of the measure to identify outlying performance. By reporting a measure mean (benchmark value), this provides an opportunity for outlying facilities to identify underperformance related to implementing applicable elements of clinical care in patients with severe sepsis and septic shock.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
Note: *Applies to all component measures, unless already endorsed or are being submitted for individual endorsement.*
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 – this measure only uses one set of specifications

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 – this measure only uses one set of specifications

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 – this measure only uses one set of specifications

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS
Note: *Applies to the overall composite measure.*

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*)

N/A – Missing data is not a concern for this measure. Please see rationale in 3c.1 in the measure submission form.

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*)

N/A

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

2c. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH
Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

See section 1b2. – analysis demonstrating the contribution of each component to the composite score

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc.*; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

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.**

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

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**.
Some data elements are in defined fields in electronic sources

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).
The SEP-1 measure is a complex composite measure which requires data abstractors to comb through documentation and interpret it in order to populate discrete fields. Currently, the narrative form of the SEP-1 measure does not make it feasible to be captured electronically. Preliminary efforts to convert SEP-1 to an eCQM within the current HQMF/QDM frameworks showed that the transition is not feasible. An example of said challenges includes the exclusion criteria for the organ dysfunction criteria in order to demonstrate the presence of severe sepsis in a patient; in order to meet exclusion criteria, abstractors have to find and interpret (difficult to find) clinician documentation which can sometimes also be written in shorthand. There are no plans to develop an eMeasure at this point.

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.
Missing data is not a concern for the SEP-1 measure as the algorithm rejects cases and does not allow submission in instances where there is missing data for a data element.
Difficulties regarding data collection and the assessment of said difficulties is one of the main driving forces for making updates to the measure. Since its inception, the measure has gone through three rounds of updates. In every single round, certain updates were made to ease abstractor burden and mitigate issues related to availability of data (vital signs for vital signs review), missing data, and frequency of data collection (Clinician attestation to suffice requirements for five data elements). All of these relevant updates have positively impacted the time and cost of data collection for providers submitting data.

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)*.
All measures which are part of CMS reporting programs are required to allow its users to not incur any costs or meet any requirements to use any aspect of the measure. All programs and tools used for the measure are required to be Open Source and free to use.

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 Payment Program Quality Improvement (external benchmarking to organizations)	Regulatory and Accreditation Programs Hospital IQR https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1138115987129

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

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

- Name of program and sponsor: Centers for Medicare & Medicaid Services
- Purpose: Hospital Inpatient Quality Reporting (IQR) Program
- Geographic area and number and percentage of accountable entities and patients included: Geographic location is nation-wide. Across the three quarters of available data, between 3,134 and 3,193 providers submitted data, which represents more than 95% of eligible providers nationwide.
- Level of measurement and setting: Acute care hospital facility level

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?*)

Measure is not currently publicly reported, but will be added to the Hospital Compare website at a date to be determined. Due to the complexity of the measure specifications, CMS desires to review and analyze the data prior to making it publicly available. There were also several updates to the specifications based on stakeholder feedback, and CMS wanted to assure stability of the specifications before public reporting.

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.*)

Measure currently used in the Hospital IQR program, with plans to add to Hospital Compare in the future to be determined.

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.

Please refer to the narrative of the results for discussion of progress on improvements across three data periods. Please refer to the table appendix for detailed tables.

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.**
N/A – None were noted

4c.2. **Please explain any unexpected benefits from implementation of this measure.**
N/A – None were noted

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.
N/A

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.**
N/A

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.
N/A

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

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

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.**
N/A

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.
No

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

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

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?

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

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.)

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 **Attachment:** v5.2a_DATAdictionary_SEP1.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Henry Ford Hospital
Co.2 **Point of Contact:** Emanuel, Rivers, erivers1@hfhs.org, 313-207-1831-
Co.3 **Measure Developer if different from Measure Steward:** Henry Ford Hospital
Co.4 **Point of Contact:** Emanuel, Rivers, erivers1@hfhs.org, 313-207-1831-

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.
1. Emmanuel Rivers, MD, MPH, FACEP, Emergency Medicine and Surgical Critical Care, Henry Ford Hospital, Institute of Medicine Fellow: measure developer, measure steward, review of current evidence, validty, reliability, usability, feasibilty, and update of measure
2. Sean R. Townsend, MD, Institute for Healthcare Improvement (IHI), California Pacific Medical Center, San Fransisco: review of current evidence, validty, reliability, usability, feasibilty, and update of measure

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2008
Ad.3 **Month and Year of most recent revision:** 06, 2012
Ad.4 **What is your frequency for review/update of this measure?** Annually for minor changes, every three years detailed review of evidence and test results.
Ad.5 **When is the next scheduled review/update for this measure?** 02, 2017

Ad.6 **Copyright statement:**
Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**