

[Click here for Pre-Evaluation Public Comments](#)[Click here for Measure Specifications](#)

Brief Measure Information

CBE #: 3752e

Corresponding Measures:

Measure Title: HIV Annual Retention in Care

Measure Steward: Health Resources and Services Administration - HIV/AIDS Bureau

sp.02. Brief Description of Measure: Percentage of patients, regardless of age, with a diagnosis of HIV who had at least two eligible encounters or at least one eligible encounter and one HIV viral load test that were at least 90 days apart within the measurement period.

1b.01. Developer Rationale: The HIV "continuum of care" is the process of HIV testing, linkage to HIV care, initiation of ART, adherence to treatment, retention in care, and virologic suppression (Gardner et al 2011). Poor retention in care is associated with lower rates of ART use (Giordano et al 2003), delayed viral suppression (Crawford et al 2014), and increased risk of mortality (Giordano et al 2007; Mugavero et al 2009). This measure will help providers direct their attention and quality improvement efforts towards improving retention in care.

CITATIONS:

Crawford, T. N., Sanderson, W. T., & Thornton, A. (2014). Impact of Poor Retention in HIV Medical Care on Time to Viral Load Suppression. *Journal of the International Association of Providers of AIDS Care (JIAPAC)*, 13(3), 242–249. <https://doi.org/10.1177/2325957413491431>.

Gardner, E. M., McLees, M. P., Steiner, J. F., Del Rio, C., & Burman, W. J. (2011). The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 52(6), 793–800. <https://doi.org/10.1093/cid/ciq243>.

Giordano, T. P., Gifford, A. L., White, A. C., Suarez-Almazor, M. E., Rabeneck, L., Hartman, C., Backus, L. I., Mole, L. A., & Morgan, R. O. (2007). Retention in care: A challenge to survival with HIV infection. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 44(11), 1493–1499. <https://doi.org/10.1086/516778>.

Giordano, T. P., White, A. C., Sajja, P., Graviss, E. A., Arduino, R. C., Adu-Oppong, A., Lahart, C. J., & Visnegarwala, F. (2003). Factors associated with the use of highly active antiretroviral therapy in patients newly entering care in an urban clinic. *Journal of Acquired Immune Deficiency Syndromes (JAIDS)*, 32(4), 399–405. <https://doi.org/10.1097/00126334-200304010-00009>

Measure Worksheet (MEW-PA-New)

Mugavero, M. J., Lin, H.-Y., Willig, J. H., Westfall, A. O., Ulett, K. B., Routman, J. S., Abrams, S., Raper, J. L., Saag, M. S., & Allison, J. J. (2009). Missed visits and mortality among patients establishing initial outpatient HIV treatment. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 48(2), 248–256. <https://doi.org/10.1086/595705>

sp.12. Numerator Statement: Number of patients who had at least two eligible encounters or at least one eligible encounter and one HIV viral load test at least 90 days apart within a 12-month measurement year.

sp.14. Denominator Statement: All patients, regardless of age, with a diagnosis of HIV during the first 8 months of the measurement period or before the measurement period who had at least one eligible encounter during the first 8 months of the measurement period.

sp.16. Denominator Exclusions: Not applicable

Measure Type: Process: Appropriate Use

sp.28. Data Source: Electronic Health Records

sp.07. Level of Analysis: Clinician: Individual

IF Endorsement Maintenance—Original Endorsement Date: N/A New Measure

Most Recent Endorsement Date: N/A New Measure

IF this measure is included in a composite, Composite#/title: N/A

IF this measure is paired/grouped, CBE#/title: N/A

sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A

Staff Assessment: New Measure

Criterion 1: Importance to Measure and Report
1a. Evidence. The evidence requirements for a <i>structure, process, or intermediate outcome</i> measure are that it is based on a systematic review (SR) and grading of the body of empirical evidence in which the specific focus of the evidence matches what is being measured. For measures derived from a patient report, the evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.
<p>The developer provides the following description for this measure:</p> <ul style="list-style-type: none"> • This is a new process measure at the individual clinician level that calculates the percentage of patients, regardless of age, with a diagnosis of HIV who had at least two eligible encounters or at least one eligible encounter and one HIV viral load test that were at least 90 days apart within the measurement period. • The developer provides a <u>logic model</u> that depicts structural inputs (HIV specialty clinicians, diagnostic laboratories) linked with expected activities/processes (provide ongoing clinical visits, conduct HIV viral load tests). The anticipated output of the activities is adherence to antiretroviral therapy (ART), which is linked with the short term outcome of HIV viral suppression, leading to long-term outcomes of improved health and reduced rates of HIV transmission.
<p>The developer provides the following evidence for this measure:</p> <ul style="list-style-type: none"> • SR of the evidence specific to this measure? <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No • Quality, Quantity, and Consistency of evidence provided? <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No • Evidence graded? <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
<p>Summary:</p> <ul style="list-style-type: none"> • The evidence cited refers primarily to two sets of clinical practice guidelines from the <u>Panel on Antiretroviral Guidelines for Adults and Adolescents</u> and the <u>International Association of Physicians in AIDS Care panel</u>; both panels recommend routine monitoring of retention in care for HIV treatment. Additional references are used to support the HIV continuum of care concept, and the linkages between poor care retention and lower rates of ART adherence, delayed viral suppression, and increased mortality risk.
<p>Exception to evidence</p> <ul style="list-style-type: none"> • N/A

Measure Worksheet (MEW-PA-New)

Questions for the Standing Committee:

- *What is the relationship between this measure and patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Does the International Association of Physicians in AIDS Care panel grade and recommendation justify a high evidence rating?*
- *Is the evidence directly applicable to the process of care being measured?*

Guidance From the Evidence Algorithm

Process measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: moderate; Consistency: moderate/high (Box 5) → Moderate (Box 5b) → Moderate.

The highest possible rating is high.

Preliminary rating for evidence: High Moderate Low Insufficient

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

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

Measure performance scores were calculated at the clinician level for the measure performance period from January 1, 2021, to December 31, 2021. The total sample included 188 clinicians and 3,136 unique, eligible patients, of which 48 clinicians had at least 11 eligible patients—the minimum sample size requirement outlined in the CMS cell suppression policy. The data for these 48 clinicians accounted for 3,065 unique patients. The measure scores are high overall, with less variability among clinicians with at least 11 patients. The 48 physicians with 11+ patients encompass 98% of the patients included in the sample. Representativeness of the clinicians included is evaluated by comparing this sample of clients of the Ryan White HIV/AIDS Program to the national average for patients with HIV (significance test not reported).

The size of the population at risk is not presented. No prior review is available.

Table 1b.02. Distribution of the measure performance scores in the clinician samples

Sample	Clinicians	Patients	Mean	SD	Min	p10	p20	p30	p40	p50	p60	p70	p80	p90	Max	IQR
All Clinicians	188	3,136	89.9	18.5	0.0	66.7	81.6	88.9	95.2	100.0	100.0	100.0	100.0	100.0	100.0	14.3
Clinicians with 11+ patients	48	3,065	89.7	8.2	58.3	80.4	83.6	86.7	88.4	91.4	92.5	94.1	96.6	100.0	100.0	9.5

Notes: SD=Standard deviation, Min=minimum, Max=maximum, p=percentile, IQR=interquartile range. These data reflect Ryan White HIV/AIDS Program clients, and thus reflect a sample of patients with higher rates of annual retention in care than the national population of people with HIV (HRSA 2022, CDC 2020).

Measure Worksheet (MEW-PA-New)

Disparities

Disparities data for retention in care are presented at the clinician level among the sample of 48 clinicians with at least 11 patients (sample of 3,065 patients) for the measure performance period from January 1, 2021, to December 31, 2021. Performance scores in deciles were stratified to assess whether there were disparities in retention in care by patients' age (< 50 years vs. >= 50 years), HIV transmission category (men who have sex with men [MSM], injection drug use [IDU], which includes both IDU and MSM and IDU, and other transmission), race (White vs. Black), and ethnicity (Hispanic or Latino vs. not Hispanic or Latino).

The mean clinician-level measure rate differences by group are in the expected direction for race (higher for whites), age (higher for ages 50+ years), and HIV transmission group (MSM only vs. IDU/MSM or other) but are not statistically significant. The developer mentions the following factors related to the lack of significance: the relatively high rate of retention in this sample; the small sample size (especially for the IDU+MSM group); and the lack of findings of disparities in recent studies.

Table 1b.04. Measure performance rates by population groups (age, HIV transmission, race, and ethnicity)

Patient Group	Clinicians	Patients	Mean	Std dev	Min	10%	20%	30%	40%	50%	60%	70%	80%	90%	Max	IQR
AGE<50	48	1,521	87.4	10.9	55.6	75.0	78.2	80.2	83.8	88.7	90.9	92.6	100.0	100.0	100.0	21.3
AGE>=50	48	1,544	91.5	10.4	50.0	77.5	86.4	90.4	91.3	93.9	95.8	100.0	100.0	100.0	100.0	11.7
MSM	42	1,228	91.0	10.0	57.1	77.1	84.8	88.3	89.3	91.4	95.8	100.0	100.0	100.0	100.0	13.1
IDU	41	239	84.3	25.3	0.0	50.0	75.0	80.8	87.5	100.0	100.0	100.0	100.0	100.0	100.0	21.9
Other transmission	42	1,005	90.8	9.2	68.2	76.5	83.3	89.9	90.5	92.9	94.1	99.2	100.0	100.0	100.0	15.6
Black	48	1,399	86.7	17.1	0.0	76.3	80.2	83.4	86.4	90.5	92.9	96.1	100.0	100.0	100.0	18.4
White	48	1,542	90.6	14.3	25.0	78.3	86.0	90.5	91.7	93.9	100.0	100.0	100.0	100.0	100.0	12.6
Not Hispanic or Latino	48	2,622	89.4	9.0	58.3	78.7	81.9	84.3	87.0	91.0	92.4	96.3	99.1	100.0	100.0	14.2
Hispanic or Latino	37	436	93.2	10.7	50.0	79.4	86.3	92.0	94.1	100.0	100.0	100.0	100.0	100.0	100.0	11.1

Notes: Results are for clinicians with ≥11 patients eligible for the denominator. SD=standard deviation, min=minimum, max=maximum, IQR=interquartile range

<p>Questions for the Standing Committee:</p> <ul style="list-style-type: none"> • <i>Is there a gap in care that warrants a national performance measure?</i> • <i>If limited disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?</i>
<p>Preliminary rating for opportunity for improvement: <input type="checkbox"/> High <input checked="" type="checkbox"/> Moderate <input type="checkbox"/> Low <input type="checkbox"/> Insufficient</p>
<p>Criteria 2: Scientific Acceptability of Measure Properties</p>
<p>Complex measure evaluated by the Scientific Methods Panel (SMP)? <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No</p>
<p>Evaluators: Staff/Laura Aume</p>
<p>2a. Reliability: Specifications and Testing</p>
<p>2a1. Specifications require the measure, as specified, to produce consistent (i.e., reliable) and credible (i.e., valid) results about the quality of care when implemented.</p> <ul style="list-style-type: none"> • The submitted measure specification follows established technical specifications for electronic clinical quality measures (eCQMs) (Quality Data Model [QDM], health quality measure format [HQMF], and Clinical Quality Language [CQL]) as indicated in subcriterion 2a1. • The submitted measure specification is fully represented and is not hindered by any limitations in the established technical specifications for eCQMs.
<p>2a2. Reliability testing demonstrates whether the measure data elements are repeatable and producing the same results a high proportion of the time when assessed in the same population in the same time period, and/or whether the measure score is precise enough to distinguish differences in performance across providers.</p>
<p>Specifications:</p> <ul style="list-style-type: none"> • eCQMs as specified using the latest industry-accepted eCQM technical specifications: HQMF, QDM, CQL, and value sets vetted through the National Library of Medicine’s (NLM) Value Set Authority Center (VSAC).
<p>Reliability Testing:</p> <ul style="list-style-type: none"> • Reliability testing conducted at the Accountable Entity Level: <ul style="list-style-type: none"> ○ Methods and results: <ul style="list-style-type: none"> ▪ Signal-to-noise (beta-binomial method) shows a range in reliability from 64% to 100% with a median of 97%. Less than 5% of the clinicians have a reliability less than 70% and more than 75% of the clinicians

Measure Worksheet (MEW-PA-New)

<ul style="list-style-type: none"> have a reliability greater than 90%. <ul style="list-style-type: none"> ▪ Split-half results: <ul style="list-style-type: none"> • Spearman rank-order correlation: 97.5% • Spearman-Brown correlation: 98.7% ▪ Test-retest results show an ICC of 94.6% ▪ Bootstrap resampling was also performed. Median values for each method are: <ul style="list-style-type: none"> • Spearman rank-order correlation: 98.4% • Spearman-Brown correlation: 99.1% • ICC: 97.4%
<p>Questions for the Standing Committee regarding reliability:</p> <ul style="list-style-type: none"> • <i>Do you have any concerns that the measure cannot be consistently implemented (i.e., are the measure specifications adequate)?</i>
<p>Guidance From the Reliability Algorithm Precise specifications (Box 1) -> Empirical testing reliability conducted (Box 2) -> Reliability testing conducted with computer measure scores (Box 4) -> Appropriate methods (Box 5) -> High (Box 6)</p> <p>The highest possible rating is high.</p>
<p>Preliminary rating for reliability: <input checked="" type="checkbox"/> High <input type="checkbox"/> Moderate <input type="checkbox"/> Low <input type="checkbox"/> Insufficient</p>
<p>2b. Validity: <u>Validity Testing</u>; <u>Exclusions</u>; <u>Risk Adjustment</u>; <u>Meaningful Differences</u>; <u>Comparability</u>; <u>Missing Data</u></p>
<p>2b2. Validity testing should demonstrate that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.</p>
<p>2b2-2b6. Potential threats to validity should be assessed/addressed.</p> <p>Validity Testing</p> <ul style="list-style-type: none"> • Validity testing conducted at the Patient/Encounter Level: <ul style="list-style-type: none"> ○ The developer assessed percent agreement and Gwet's AC1 for comparison between EHR data and manual abstraction ○ There developer assessed percent agreement and Gwet's AC1 at least 99% for all elements except HIV diagnosis date (Percent agreement: 75%, Gwet's AC1: 68%) ○ Although there is some disagreement for HIV diagnosis date it does not affect inclusion in the denominator for any of the patients.

Measure Worksheet (MEW-PA-New)

- Validity testing conducted at the Accountable Entity Level:
 - Difference in performance scores between groups known to differ in quality performed:
 - Difference between Age < 50 years and Age ≥ 50 years is not statistically significant with 95% confidence (p=0.065).
 - Difference and between transmission category (men who have sex with men [MSM] versus non-MSM) is not statistically significant with 95% confidence (p=0.100).
 - Face Validity assessment shows 71% of clinicians support the measure's numerator and denominator specifications, and 57% of clinicians agree that the measure can distinguish quality of care. 88% of the Technical Expert Panel agree that the measure is important and related to quality of care.
- Feasibility testing was conducted at 7 test sites, which includes 4 different EHR systems.
- The Feasibility Scorecard indicated that the following data elements have issues with accuracy: (note: these data elements were not available at one or more sites, and as a result scored low on accuracy as well. Neither is required to calculate the measure score):
 - "Encounter Performed: Home Healthcare Services"
 - "Encounter, Performed: Outpatient Consultation"
- Data elements:
 - "Diagnosis: HIV"
 - "Encounter, Performed: Face-to-Face Interaction"
 - "Encounter, Performed: Office Visit"
 - "Encounter, Performed: Outpatient Consultation"
 - "Encounter, Performed: Preventive Care - Established Office Visit, 0 to 17"
 - "Encounter, Performed: Preventive Care Services - Established Office Visit, 18 and Up"
 - "Encounter, Performed: Preventive Care Services-Initial Office Visit, 18 and Up"
 - "Encounter, Performed: Preventive Care- Initial Office Visit, 0 to 17"
 - "Encounter, Performed: Telehealth Services"
 - "Encounter, Performed: Telephone Visits"
 - "Encounter Performed: Annual Wellness Visit"
 - "Encounter Performed: Home Healthcare Services"
 - "Encounter Performed: Preventive Care Services Other"
 - "Laboratory Test, Performed: HIV Viral Load"
 - "Patient Characteristic Ethnicity: Ethnicity"
 - "Patient Characteristic Payer: Payer"
 - "Patient Characteristic Race: Race"
 - "Patient Characteristic Sex: ONC Administrative Sex"

Measure Worksheet (MEW-PA-New)

Exclusions <ul style="list-style-type: none"> The measure does not use exclusions.
Risk Adjustment <ul style="list-style-type: none"> The measure is not risk-adjusted or stratified.
Meaningful Differences <ul style="list-style-type: none"> 10% of the clinicians perform statistically significantly lower and 23% of the clinicians perform statistically significantly higher than average.
Missing Data <ul style="list-style-type: none"> The measure is based on the presence of encounters, tests, and HIV diagnosis dates, so missing data are not a threat to validity.
Comparability <ul style="list-style-type: none"> The measure only uses one set of specifications for this measure.
Questions for the Standing Committee regarding validity: <ul style="list-style-type: none"> <i>Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk adjustment approach, etc.)?</i> <i>Are the accuracy issues that are captured in the Feasibility Scorecard substantial enough to impact the validity of these data elements?</i>
Guidance From the Validity Algorithm All threats assessed (Box 1) -> Empirical validity testing conducted on the measure as specified (Box 2) -> Validity testing conducted with computer measure scores (Box 5) -> Appropriate methods (Box 6) -> Moderate (Box 7) The highest possible rating is high.
Preliminary rating for validity: <input type="checkbox"/> High <input checked="" type="checkbox"/> Moderate <input type="checkbox"/> Low <input type="checkbox"/> Insufficient
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.

Measure Worksheet (MEW-PA-New)

- This measure is generated or collected by and used by healthcare personnel during the provision of care.
- All data elements are in defined fields in electronic claims.
- There are no fees or licenses required for usage of this measure.
- Using a simulated data set, the submission demonstrates that the evaluation of 100 percent of the measure logic can be automated.
- The required data elements are widely available in electronic health data. The two encounter type data elements that were missing from test sites ("Encounter Performed: Home Healthcare Services" and "Encounter, Performed: Outpatient Consultation") are (1) not required for measure score calculation, given the availability of other eligible encounter types and (2) only missing because they are not applicable to the test sites. "Diagnosis: HIV" was not available in a structured data field at 2 of the 7 test sites, although both collect this information in unstructured formats. One of these sites already has a plan in place to change workflows to capture this information in a structured field. Given the large share of test sites that captured this information or expect to do so in the future, measure data element feasibility is moderate.

Questions for the Standing 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?*
- *For data elements assessed to have feasibility issues, does the developer present a credible, near-term path to electronic collection?*

Preliminary rating for feasibility: High Moderate Low Insufficient

Criterion 4: Use and Usability
4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

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

4a1. 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 they are not in use at the time of initial endorsement, then a credible plan for implementation within the specified time frames is provided.

Measure Worksheet (MEW-PA-New)

Current uses of the measure Publicly reported? <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No Current use in an accountability program? <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No <input type="checkbox"/> UNCLEAR Planned use in an accountability program? <input checked="" type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> N/A			
Accountability program details <ul style="list-style-type: none"> HRSA plans to submit this measure for use as a clinician-level measure in the CMS Merit Based Incentive Program (MIPS) program. HRSA will submit the measure to the 2023 Measures Under Consideration (MUC) list by April 30, 2023. 			
4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: (1) Those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; (2) Those being measured, and other users have been given an opportunity to provide feedback on the measure performance or implementation; and (3) This feedback has been considered when changes are incorporated into the measure.			
Feedback on the measure provided by those being measured or others <ul style="list-style-type: none"> During the testing process, the developer provided clinical practices that participated in testing with measure specifications, data dictionaries, value sets, and fact sheets to assist them in generating datasets used for testing. This measure has not yet been implemented and additional feedback has not yet been provided. 			
Questions for the Standing Committee: <ul style="list-style-type: none"> <i>How have (or can) the performance results be used to further the goal of high quality, efficient healthcare?</i> <i>How has the measure been vetted in real-world settings by those being measured or others?</i> 			
Preliminary rating for Use: <input checked="" type="checkbox"/> Pass <input type="checkbox"/> No Pass			
4b. Usability (4b1. Improvement; 4b2. Benefits of measure)			
4b. Usability evaluates the extent to which audiences (e.g., consumers, purchasers, providers, and policymakers) use or could use performance results for both accountability and performance improvement activities.			
4b1 Improvement. Progress toward achieving the goal of high quality, efficient healthcare for individuals or populations is demonstrated.			
Improvement results <ul style="list-style-type: none"> This measure is not currently in use as an eCQM in a quality improvement program. Data to support progress on improvement were not provided. 			

Measure Worksheet (MEW-PA-New)

4b2. Benefits versus harms. 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).

Unexpected findings (positive or negative) during implementation

- This measure has not yet been implemented.

Potential harms

- Information on potential harms was not provided by the developer.

Additional Feedback:

- N/A

Questions for the Standing 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?*

Preliminary rating for Usability and Use:

High Moderate Low Insufficient

Criterion 5: Related and Competing Measures

Related Measures

- 3210e: HIV Viral Load Suppression
- 3211e: Prescription of HIV Antiretroviral Therapy
- 0409: HIV/AIDS: Sexually Transmitted Diseases- Screening for Chlamydia, Gonorrhea, and Syphilis (CQM only)
- 3209e: HIV Medical Visit Frequency

Measure Worksheet (MEW-PA-New)

Harmonization

- The denominator population for this measure differs slightly from three related measures—3210e, 3211e, and 0409—with respect to the timing of the patient’s HIV diagnosis and eligible encounter, and these differences are due to the specific timing required for assessing retention in care.
- The population is limited to patients diagnosed no earlier than eight months into the performance period, and with an eligible encounter no later than eight months into the performance period, to allow sufficient time for a clinician to have a follow-up encounter or viral load test 90 days after the initial encounter and before the end of the measurement period.
- The measure also differs from the denominator population for 0409 with respect to the patient’s age because according to the developer, retention in care is a relevant process measure for all patients with HIV, regardless of age, while 0409 focuses on older patients who may be sexually active.
- This measure and 3209e use different approaches for measure retention in care for people with HIV. This measure defines retention based on encounters and viral load tests within a one-year period, assessing whether patients had at least two encounters or one encounter and one viral load at least 90 days apart. 3209e assesses retention over a two-year period, assessing whether patients had a medical visit at least once every six months over that period.

QUALITY MEASURE SUBMISSION FORM

Version: 1.0; Generated: 13 April 2023

Introduction

Thank you for your interest in submitting a measure to Battelle for possible endorsement.

What criteria are used to evaluate measures? Measures are evaluated on standardized criteria: importance to measure and report, scientific acceptability of measure properties, feasibility, usability and use, and related and competing measures. For your measure to be evaluated against these measure evaluation criteria, you must complete the measure submission form.

Why do I have to complete a form? Due to the volume and/or complexity of proposed measures, Battelle provides measure information to committee reviewers in a standardized format to facilitate their evaluation of whether the measure meets the measure evaluation criteria. This form allows the measure steward to present information demonstrating that the proposed measure meets endorsement criteria.

What is on the form? The information requested in this form is directly related to the measure evaluation criteria.

Can't I just submit our files for consideration? No. Measures must be submitted through the online form to be considered for the Spring 2023 cycle. Requested information should be entered directly into this form and as well as any necessary or required attachments.

Can I submit additional details and materials? Additional materials will be considered only as supplemental. Do NOT rely on material provided in an appendix to provide measure specifications or to demonstrate meeting the criteria. The core information needed to evaluate the measure should be provided in the appropriate submission form fields and required attachments. Please contact PQMsupport@battelle.org regarding questions about submitting supplemental materials.

What do I do first? If you have started a new submission by answering five qualifying questions, you may proceed to the "Previous Submission Information" tab to continue with your submission. The "Conditions" tab will list the conditions that must be met before your proposed measures may be considered and evaluated for suitability as endorsed voluntary consensus standards. You are asked to acknowledge reading and accepting the conditions.



Can I make changes to a form once I have submitted it? No. Once you submit your measure, you will NOT be able to return to this submission form to make further revisions. You will need to contact project staff.

What if I need additional help? Please contact the project staff at PQMsupport@battelle.org if you have questions regarding the information requested or submitting supplemental materials.

NOTE: All measure submissions should be 508-compliant. Refer to the Checklist for Developer 508 Guidelines (PDF) to ensure all guidelines apply to all parts of your submission, including all fields and attachments used within the measure submission form.

Please email us at PQMsupport@battelle.org if you experience technical difficulties using the online submission form.

Thank you for your interest in submitting measures to Battelle.

Previous Submission Information (1 – 4)

1) Select whether this measure was previously submitted to the prior consensus-based entity (the National Quality Forum [NQF]) and given an identifying number.

- Previously submitted to NQF
- New measure, never submitted.

2) Provide the measure number of the previously submitted measure.

Not applicable.

3) If the measure has an electronic clinical quality measure (eCQM) version, provide the measure number of the previously submitted measure.

Not applicable.

4) If this eCQM has a registry version, provide the measure numbers of the previously submitted measure.

Not applicable.

Conditions (1 - 2)

Several conditions must be met before proposed measures may be considered and evaluated for suitability as voluntary consensus standards. If any of the conditions are not met, the measure will not be accepted for consideration.

- A. A Measure Steward Agreement is signed or the steward is a government organization. (All non-government organizations must sign a Measure Steward Agreement.) For more information about completing a Measure Steward Agreement, please go to: [Endorsement | Partnership for Quality Measurement \(p4qm.org\)](https://p4qm.org) and follow the instructions.
- B. The measure owner/steward verifies there is an identified responsible entity and a process to maintain and update the measure on a schedule that is commensurate with the rate of clinical innovation, but at least every three years.
- C. The intended use of the measure includes both accountability applications (including public reporting) and performance improvement to achieve high-quality, efficient healthcare.
- D. The measure is fully specified and tested for reliability and validity.
- E. The measure developer/steward attests that harmonization with related measures and issues with competing measures have been considered and addressed, as appropriate.
- F. The requested measure submission information is complete and responsive to the questions so that all the information needed to evaluate all criteria is provided.

1) Check if either of the following apply.

- Proprietary measure or components (e.g., risk model, codes)
- Proprietary measure or components with fees
- None of the above

2) Check the box below to agree to the conditions listed above.

- I have read and accept the conditions as specified above

Specifications: Maintenance Update (spma.01 - spma.02)

spma.01) Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.

- No
 Yes

spma.02) Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

For example, specifications may have been updated based on suggestions from a previous measure endorsement review.

Not applicable.

Measure Specifications (sp.01 - sp.32)

sp.01) Provide the measure title.

Measure titles should be concise yet convey who and what is being measured.

HIV Annual Retention in Care

sp.02) Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

Percentage of patients, regardless of age, with a diagnosis of HIV who had at least two eligible encounters or at least one eligible encounter and one HIV viral load test that were at least 90 days apart within the measurement period

sp.03) Provide a rationale for why this measure must be reported with other measures to appropriately interpret results.

Not applicable.

sp.04) Check all the clinical condition/topic areas that apply to your measure, below.

- Behavioral Health
- Behavioral Health: Alcohol, Substance Use/Abuse
- Behavioral Health: Anxiety
- Behavioral Health: Attention Deficit Hyperactivity Disorder (ADHD)
- Behavioral Health: Bipolar Disorder
- Behavioral Health: Depression
- Behavioral Health: Domestic Violence
- Behavioral Health: Other Serious Mental Illness
- Behavioral Health: Post-Traumatic Stress Disorder (PTSD)
- Behavioral Health: Schizophrenia
- Behavioral Health: Suicide
- Cancer
- Cancer: Bladder
- Cancer: Breast
- Cancer: Colorectal
- Cancer: Gynecologic
- Cancer: Hematologic
- Cancer: Liver

- Cancer: Lung, Esophageal
- Cancer: Prostate
- Cancer: Renal
- Cancer: Skin
- Cancer: Thyroid
- Cardiovascular
- Cardiovascular: Arrhythmia
- Cardiovascular: Congestive Heart Failure
- Cardiovascular: Coronary Artery Disease
- Cardiovascular: Coronary Artery Disease (AMI)
- Cardiovascular: Coronary Artery Disease (PCI)
- Cardiovascular: Hyperlipidemia
- Cardiovascular: Hypertension
- Cardiovascular: Secondary Prevention
- Critical Care
- Critical Care: Assisted Ventilation
- Critical Care: Intensive Monitoring
- Dental
- Dental: Caries
- Dental: Tooth Loss
- Ears, Nose, Throat (ENT)
- Ears, Nose, Throat (ENT): Ear Infection
- Ears, Nose, Throat (ENT): Hearing
- Ears, Nose, Throat (ENT): Pharyngitis
- Ears, Nose, Throat (ENT): Tonsillitis
- Endocrine
- Endocrine: Calcium and Metabolic Bone Disorders
- Endocrine: Diabetes
- Endocrine: Female and Male Endocrine Disorders
- Endocrine: Hypothalamic-Pituitary Disorders
- Endocrine: Thyroid Disorders
- Eye Care
- Eye Care: Age-related macular degeneration (AMD)
- Eye Care: Cataracts
- Eye Care: Diabetic retinopathy
- Eye Care: Glaucoma
- Gastrointestinal (GI)
- Gastrointestinal (GI): Constipation
- Gastrointestinal (GI): Gall Bladder Disease
- Gastrointestinal (GI): Gastroenteritis

- Gastrointestinal (GI): Gastro-Esophageal Reflux Disease (GERD)
- Gastrointestinal (GI): Hemorrhoids
- Gastrointestinal (GI): Hernia
- Gastrointestinal (GI): Inflammatory Bowel Disease
- Gastrointestinal (GI): Irritable Bowel Syndrome
- Gastrointestinal (GI): Peptic Ulcer
- Genitourinary (GU)
- Genitourinary (GU): Benign Prostatic Hyperplasia
- Genitourinary (GU): Erectile Dysfunction/Premature Ejaculation
- Genitourinary (GU): Incontinence/pelvic floor disorders
- Genitourinary (GU): Prostatitis
- Genitourinary (GU): Urinary Tract Infection (UTI)
- Gynecology (GYN)
- Gynecology (GYN): Abnormal bleeding
- Gynecology (GYN): Endometriosis
- Gynecology (GYN): Infections
- Gynecology (GYN): Menopause
- Gynecology (GYN): Pelvic Pain
- Gynecology (GYN): Uterine fibroids
- Infectious Diseases (ID)
- Infectious Diseases (ID): HIV/AIDS
- Infectious Diseases (ID): Influenza
- Infectious Diseases (ID): Lyme Disease
- Infectious Diseases (ID): Meningococcal Disease
- Infectious Diseases (ID): Pneumonia and respiratory infections
- Infectious Diseases (ID): Sepsis
- Infectious Diseases (ID): Sexually Transmitted
- Infectious Diseases (ID): Tuberculosis
- Liver
- Liver: Viral Hepatitis
- Musculoskeletal
- Musculoskeletal: Falls and Traumatic Injury
- Musculoskeletal: Gout
- Musculoskeletal: Joint Surgery
- Musculoskeletal: Low Back Pain
- Musculoskeletal: Osteoarthritis
- Musculoskeletal: Osteoporosis
- Musculoskeletal: Rheumatoid Arthritis
- Neurology
- Neurology: Alzheimer's Disease

- Neurology: Autism
- Neurology: Brain Injury
- Neurology: Epilepsy
- Neurology: Migraine
- Neurology: Parkinson's Disease
- Neurology: Spinal Cord Injury
- Neurology: Stroke/Transient Ischemic Attack (TIA)
- Other (please specify here:)
- Palliative Care and End-of-Life Care
- Palliative Care and End-of-Life Care: Advanced Directives
- Palliative Care and End-of-Life Care: Amyotrophic Lateral Sclerosis (ALS)
- Palliative Care and End-of-Life Care: Hospice Management
- Palliative Care and End-of-Life Care: Inappropriate use of acute care services
- Palliative Care and End-of-Life Care: Pain Management
- Perinatal Health
- Perinatal Health: Labor and Delivery
- Perinatal Health: Newborn Care
- Perinatal Health: Post-Partum Care
- Perinatal Health: Preconception Care
- Perinatal Health: Prenatal Care
- Renal
- Renal: Acute Kidney Injury
- Renal: Chronic Kidney Disease (CKD)
- Renal: End Stage Renal Disease (ESRD)
- Renal: Infections
- Reproductive Health
- Reproductive Health: Family planning and contraception
- Reproductive Health: Infertility
- Reproductive Health: Male reproductive health
- Respiratory
- Respiratory: Acute Bronchitis
- Respiratory: Allergy
- Respiratory: Asthma
- Respiratory: Chronic Obstructive Pulmonary Disease (COPD)
- Respiratory: Dyspnea
- Respiratory: Pneumonia
- Respiratory: Sleep Apnea
- Surgery
- Surgery: Cardiac Surgery
- Surgery: Colorectal

- Surgery: Neurosurgery / Spinal
- Surgery: Orthopedic
- Surgery: Orthopedic Hip/Pelvic Fractures
- Surgery: Pediatric
- Surgery: Perioperative and Anesthesia
- Surgery: Plastic
- Surgery: Thoracic Surgery
- Surgery: Trauma
- Surgery: Vascular Surgery

sp.05) Check all the non-condition specific measure domain areas that apply to your measure, below.

- Access to Care
- Care Coordination
- Care Coordination: Readmissions
- Care Coordination: Transitions of Care
- Disparities Sensitive
- Health and Functional Status
- Health and Functional Status: Change
- Health and Functional Status: Nutrition
- Health and Functional Status: Obesity
- Health and Functional Status: Physical Activity
- Health and Functional Status: Quality of Life
- Health and Functional Status: Total Health
- Immunization
- Other (please specify here:)
- Person-and Family-Centered Care: Person-and Family-Centered Care
- Person-and Family-Centered Care: Workforce
- Primary Prevention
- Primary Prevention: Nutrition
- Primary Prevention: Tobacco Use
- Safety
- Safety: Complications
- Safety: Healthcare Associated Infections
- Safety: Medication
- Safety: Overuse
- Screening

sp.06) Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

- Adults (Age >= 18)
- Children (Age < 18)
- Elderly (Age >= 65)
- Populations at Risk: Dual eligible beneficiaries of Medicare and Medicaid
- Populations at Risk: Individuals with multiple chronic conditions
- Populations at Risk: Veterans
- Women

sp.07) Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

- Accountable Care Organization
- Clinician: Group/Practice
- Clinician: Individual
- Facility
- Health Plan
- Integrated Delivery System
- Other (please specify here:)
- Population: Community, County or City
- Population: Regional and State

sp.08) Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.

- Ambulatory Care
- Behavioral Health
- Home Care
- Inpatient/Hospital
- Other (please specify here:)
- Outpatient Services
- Post-Acute Care

sp.09) Provide a Uniform Resource Locator (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. If no URL is available, indicate "none available".

None available.

sp.10) Indicate whether Health Quality Measure Format (HQMF) specifications are attached.

Attach the zipped output from the measure authoring tool (MAT) for eCQMs - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications). HQMF specifications are attached.

HQMF specifications are NOT attached (Please explain).

sp.11) Attach the simulated testing attachment.

All eCQMs require a simulated testing attachment to confirm that the HTML output from Bonnie testing (or testing of some other simulated data set) includes 100% coverage of measured patient population testing, with pass/fail test cases for each sub-population. This can be submitted in the form of a screenshot.

Testing is attached

Testing is NOT attached (please explain)

sp.12) Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff at PQMsupport@battelle.org. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

Available in attached Excel or csv file

No data dictionary/code table – all information provided in the submission form

For the question below: state the outcome/process being measured. Calculations of the risk-adjusted outcome measures should be described in sp.22.

sp.13) State the numerator.

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.

Number of patients who had at least two eligible encounters or at least one eligible encounter and one HIV viral load test at least 90 days apart within a 12-month measurement year

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.14) Provide details needed to calculate the numerator.

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 sp.11.

The list of qualifying encounters and viral load tests are in the attached file (see also value sets in sp.12 and specifications in sp.10).

Measurement period is equivalent to a calendar year.

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.15) State the denominator.

Brief, narrative description of the target population being measured.

All patients, regardless of age, with a diagnosis of HIV during the first 8 months of the measurement period or before the measurement period who had at least one eligible encounter during the first 8 months of the measurement period

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.16) Provide details needed to calculate the denominator.

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 sp.11.

The list of qualifying HIV diagnoses and eligible encounters are in the attached file (see also value sets in sp.12 and specifications in sp.10).

Patient age

HIV diagnosis date

Note that the requirement for the eligible encounter to occur within the first eight months

of the measurement period is to allow for sufficient time to complete the follow-up encounter or laboratory visit that would meet the requirement for at least 90 days in between encounters. As an example, if a patient has an eligible encounter on August 31st, then the first possible date of a potential follow-up encounter would be November 29th. By restricting the denominator to eligible encounters through the first eight months, then the clinician would have approximately 1 month to complete a follow-up encounter.

sp.17) Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

Not applicable.

sp.18) Provide details needed to calculate the denominator exclusions.

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 sp.11.

Not applicable.

sp.19) Provide all information required to stratify the measure results, if necessary.

Include 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 in the Data Dictionary field.

Not applicable.

sp.20) Is this measure adjusted for socioeconomic status (SES)?

- Yes
- No

sp.21) Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

- No risk adjustment or risk stratification
- Statistical risk model

- Stratification by risk category/subgroup (specify number of risk factors)
- Other approach to address risk factors (please specify here:)

sp.22) Select the most relevant type of score.

Attachment: If available, please provide a sample report.

- Categorical, e.g., yes/no
- Continuous variable, e.g. average
- Count
- Frequency Distribution
- Non-weighted score/composite/scale
- Other (please specify here:)
- Rate/proportion
- Ratio
- Weighted score/composite scale

sp.23) Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score.

- Better quality = Higher score
- Better quality = Lower score
- Better quality = Score within a defined interval
- Passing score defines better quality

sp.24) Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

Denominator

1. Identify patients with an eligible encounter in the first eight months of the measurement period
2. Restrict to all patients diagnosed with HIV during the first eight months of the measurement period or any time prior.
3. Patients meeting these criteria are in the denominator.

Numerator

1. Identify denominator eligible patients who had at least (a) two eligible encounters or (b) one eligible encounter and one viral load test at least 90 days apart
2. Patients meeting either criterion in step one are included in the numerator, and patients that do not are not in the numerator.

sp.25) Attach a copy of the instrument (e.g. survey, tool, questionnaire, scale) used as a data source for your measure, if available.

- Copy of instrument is attached.
- Copy of instrument is NOT attached (please explain).

The measure utilizes structured fields from electronic health record (EHR) data.

sp.26) Indicate the responder for your instrument.

- Patient
- Family or other caregiver
- Clinician
- Other (specify)

Not applicable; not a survey-based measure.

sp.27) If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

- *Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.*
- *The sample should represent the variety of entities whose performance will be measured. The samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.*
- *The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.*
- *When possible, units of measurement and patients within units should be randomly selected.*

We recruited clinicians from 7 test sites that provide clinical care to patients with HIV and receive funding from the Ryan White HIV/AIDS Program. The characteristics of the

test sites are listed below. These sites represented different regions, covering both urban and rural areas, and using different EHR systems to ensure the generalizability of findings. All clinicians that have at least 11 patients eligible for the measure denominator are included in the analysis for a total of 48 clinicians and 3,065 patients. The patients included in the analysis are Ryan White HIV/AIDS Program recipients that had an eligible encounter within the measurement period (January 1, 2021 to December 31, 2021) for those 48 clinicians. As a part of reliability testing in order to add rigor to the limited number of unique clinicians, we also conducted bootstrap resampling. This approach was used to test the stability of the measure rates over 2,000 replications of the initial sample. Results from the bootstrap testing (section 2a.11) support the generalizability of the findings.

/ **Geographic region and urban/rural communities served**

- Four Northeast; two Midwest; one South
- Five urban; two combination of urban and rural

/ **Clinic types**

- Two hospital or university-based clinics
- Four publicly funded community health centers
- One other community-based service organization

/ **Electronic health record (EHR)**

- eClinicalWorks (3)
- EPIC (2)
- NextGen (1)
- Athena Health (1)

sp.28) Identify whether and how proxy responses are allowed.

Not applicable.

sp.29) Survey/Patient-reported data.

Provide instructions for data collection and guidance on minimum response rate. Specify calculation of response rates to be reported with performance measure results.

Not applicable.

sp.30) Select only the data sources for which the measure is specified.

- Assessment Data
- Claims
- Electronic Health Data
- Electronic Health Records
- Instrument-Based Data
- Management Data

- Other (please specify here:)
- Paper Medical Records
- Registry Data

sp.31) Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

The measure is calculated based on structured data pulled from each submitter's electronic health record.

sp.32) Provide the data collection instrument.

- Available at measure-specific web page URL identified in sp.09
- Available in attached appendix in Question 1 of the Additional Section
- No data collection instrument provided

Importance to Measure and Report: Maintenance of Endorsement (1ma.01)

1ma.01) Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.

- Yes
- No

Importance to Measure and Report: Evidence (Complete for Outcome Measures) (1a.01 - 1a.03)

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01) Provide a logic model.

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.

1a.02) Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.

Describe how and from whom input was obtained.

1a.03) Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

Importance to Measure and Report: Evidence (Complete for Process Measures) (1a.03 - 1a.16)

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

Current Submission:

Updated evidence information here.

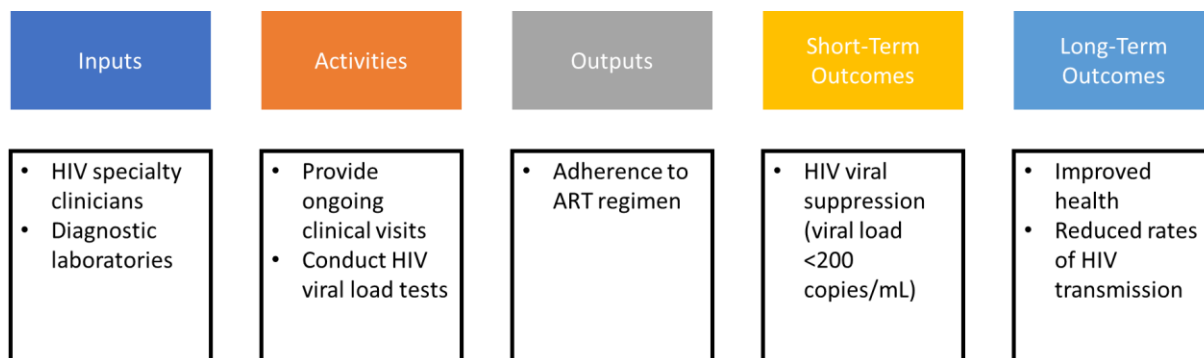
Previous (Year) Submission:

Evidence from the previous submission here.

1a.01) Provide a logic model.

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.

Figure 1a.01 Logic Model



The HIV "continuum of care" is the process of HIV testing, linkage to HIV care, initiation of ART, adherence to treatment, retention in care, and virologic suppression (Gardner et al 2011). Poor retention in care is associated with lower rates of ART use (Giordano et al 2003), delayed viral suppression (Crawford et al 2014), and increased risk of mortality (Giordano et al 2007; Mugavero et al 2009). Inputs to the process include HIV specialty clinicians and diagnostic laboratories. These inputs feed into the following activities: HIV specialty clinicians provide ongoing clinic visits; and HIV specialty clinicians refer their patients to diagnostic laboratories, which conduct HIV viral load tests. These activities result in the output of patient adherence to their ART regimen (Meyers et al., 2019). This output results in the short-term outcome of HIV viral suppression, defined as a viral load < 200 copies/mL (Byrd et al., 2019). This short-term outcome leads to the longer-

term outcomes of improved health and reduced rates of HIV transmission (Cohen et al., 2011).

ALT-TEXT:

Figure 1a.01 shows the inputs, activities, outputs, short-term outcomes, and long-term outcomes involved in the “HIV continuum of care”. The inputs include HIV specialty clinicians and diagnostic laboratories. The activities include providing ongoing clinical visits and conducting HIV viral load tests, and the output is adherence to ART. The short-term outcomes are HIV viral suppression, and the long-term outcomes are improved health and reduced rates of HIV transmission.

CITATIONS:

Byrd, K. K., Hou, J. G., Hazen, R., Kirkham, H., Suzuki, S., Clay, P. G., Bush, T., Camp, N. M., Weidle, P. J., Delpino, A., & Patient-Centered HIV Care Model Team (2019). *Antiretroviral Adherence Level Necessary for HIV Viral Suppression Using Real-World Data*. *Journal of acquired immune deficiency syndromes* (1999), 82(3), 245–251. <https://doi.org/10.1097/QAI.0000000000002142>

Cohen, M. S., Chen, Y. Q., McCauley, M., Gamble, T., Hosseinipour, M. C., Kumarasamy, N., Hakim, J. G., Kumwenda, J., Grinsztejn, B., Pilotto, J. H., Godbole, S. V., Mehendale, S., Chariyalertsak, S., Santos, B. R., Mayer, K. H., Hoffman, I. F., Eshleman, S. H., Piwowar-Manning, E., Wang, L., Makhema, J., ... HPTN 052 Study Team (2011). *Prevention of HIV-1 infection with early antiretroviral therapy*. *The New England journal of medicine*, 365(6), 493–505. <https://doi.org/10.1056/NEJMoa1105243>

Crawford, T. N., Sanderson, W. T., & Thornton, A. (2014). *Impact of Poor Retention in HIV Medical Care on Time to Viral Load Suppression*. *Journal of the International Association of Providers of AIDS Care (JIAPAC)*, 13(3), 242–249. <https://doi.org/10.1177/2325957413491431>.

Gardner, E. M., McLees, M. P., Steiner, J. F., Del Rio, C., & Burman, W. J. (2011). *The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection*. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 52(6), 793–800. <https://doi.org/10.1093/cid/ciq243>.

Giordano, T. P., Gifford, A. L., White, A. C., Suarez-Almazor, M. E., Rabeneck, L., Hartman, C., Backus, L. I., Mole, L. A., & Morgan, R. O. (2007). *Retention in care: A challenge to survival with HIV infection*. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 44(11), 1493–1499. <https://doi.org/10.1086/516778>.

Giordano, T. P., White, A. C., Sajja, P., Graviss, E. A., Arduino, R. C., Adu-Oppong, A., Lahart, C. J., & Visnagarwala, F. (2003). *Factors associated with the use of highly active antiretroviral therapy in patients newly entering care in an urban clinic*. *Journal of Acquired Immune Deficiency Syndromes (JAIDS)*, 32(4), 399–405.

<https://doi.org/10.1097/00126334-200304010-00009>.

Mugavero, M. J., Lin, H.-Y., Willig, J. H., Westfall, A. O., Ulett, K. B., Routman, J. S., Abrams, S., Raper, J. L., Saag, M. S., & Allison, J. J. (2009). Missed visits and mortality among patients establishing initial outpatient HIV treatment. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 48(2), 248–256. <https://doi.org/10.1086/595705>.

Meyers, D. J., Cole, M. B., Rahman, M., Lee, Y., Rogers, W., Gutman, R., & Wilson, I. B. (2019). The association of provider and practice factors with HIV antiretroviral therapy adherence. *AIDS (London, England)*, 33(13), 2081–2089. <https://doi.org/10.1097/QAD.0000000000002316>.

1a.02) Select the type of source for 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.

- 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 (please specify here:)

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, you may add additional tables to the relevant sections. Please follow the 508 Checklist for tables.

Evidence - Systematic Reviews Table (Repeatable)

1a.03) Provide the title, author, date, citation (including page number) and URL for the systematic review.

Panel on Antiretroviral Guidelines for Adults and Adolescents. “Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV.” Washington, DC: U.S. Department of Health and Human Services, 2022. Available at <https://clinicalinfo.hiv.gov/sites/default/files/guidelines/documents/adult-adolescent-arv/guidelines-adult-adolescent-arv.pdf> Updated September 21, 2022.

Thompson, M. A., Mugavero, M. J., Amico, K. R., Cargill, V. A., Chang, L. W., Gross, R., Orrell, C., Altice, F. L., Bangsberg, D. R., Bartlett, J. G., Beckwith, C. G., Dowshen,

N., Gordon, C. M., Horn, T., Kumar, P., Scott, J. D., Stirratt, M. J., Remien, R. H., Simoni, J. M., & Nachega, J. B. (2012). Guidelines for improving entry into and retention in care and antiretroviral adherence for persons with HIV: evidence-based recommendations from an International Association of Physicians in AIDS Care panel. *Annals of internal medicine*, 156(11), 817–294. <https://doi.org/10.7326/0003-4819-156-11-201206050-00419>

1a.04) Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

Retention in care should be routinely monitored. There are various ways to measure retention, including measures based on attended visits over a defined period of time (constancy measures) and measures based on missed visits. Both approaches are valid and independently predict survival. Missed visits and a prolonged time since the last visit are relatively easy to measure and should trigger efforts to retain or re-engage a person in care. Constancy measures (e.g., at least two visits that are at least 90 days apart over 1 year or at least one visit every 6 months over the last 2 years) can be used as clinic quality assurance measures.” (Panel on Antiretroviral Guidelines for Adults and Adolescents, p. L-4)

“Poor retention in HIV care is associated with greater risk of death. Poor retention is more common in people who use substances, have serious mental health problems, have unmet socioeconomic needs (e.g., housing, food, transportation), lack financial resources or health insurance, have schedules that complicate adherence, have been recently incarcerated, or face stigma. At the provider and health system level, low trust in providers and a poor patient–provider relationship have been associated with lower retention, as has lower satisfaction with the clinic experience. Availability of appointments and timeliness of appointments (i.e., long delay from the request for an appointment to the appointment’s date) and scheduling convenience are also factors.” (Panel on Antiretroviral Guidelines for Adults and Adolescents , p. L-3)

“Recommendation 2: Systematic monitoring of retention in HIV care is recommended for all patients (II A): Retention in care is associated with improved individual health outcomes, including HIV biomarker and clinical variables, and may reduce community-level viral burden, with implications for secondary prevention. Although monitoring retention is routinely recommended, specific details, such as retention measures to be used and desired visit frequency, vary among jurisdictions and programs and should be in harmony with national and international guidelines. Many retention measures (for example, visit adherence, gaps in care, and visits per interval of time) and data sources (for example, surveillance, medical records, and administrative databases) have been used and may be applied in accordance with local resources and standards of care. As with monitoring of linkage, integration of data sources may enhance monitoring of retention.” (International Association of Physicians in AIDS Care panel, page 4)

1a.05) Provide the grade assigned to the evidence associated with the

recommendation and include the definition of the grade.

Panel on Antiretroviral Guidelines for Adults and Adolescents- evidence not graded.

International Association of Physicians in AIDS Care panel- high (II)

1a.06) Provide all other grades and definitions from the evidence grading system.

Table 1a.06. International Association of Physicians in AIDS Care panel’s evidence and recommendation grading system

Category/Grade	Definition
Quality of the body of evidence	Definition
Excellent (I)	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 evidence without important limitations.
Low (IV)	Observational study evidence with important or critical limitations.
Strength of recommendation	Definition
Strong (A)	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 on the basis of individual patient circumstances. Not recommended routinely.

ALT-TEXT:

Table 1a.06 shows the International Association of Physicians in AIDS Care panel’s evidence and recommendations grading system. For Quality of the body of evidence, the category or grades and definitions are as follows: Excellent (I), 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

evidence without important limitations. Low (IV), observational study evidence with important or critical limitations. For Strength of recommendation, the category or grades and definitions are as follows: Strong (A), 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 on the basis of individual patient circumstances. Not recommended routinely.

1a.07) Provide the grade assigned to the recommendation, with definition of the grade.

Panel on Antiretroviral Guidelines for Adults and Adolescents- recommendation not graded.

International Association of Physicians in AIDS Care panel- Strong (A)

1a.08) Provide all other grades and definitions from the recommendation grading system.

See Table 1a.06

1a.09) Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

Recommendations from the Panel on Antiretroviral Guidelines for Adults and Adolescents referenced above cited 13 different studies, including 7 retrospective cohort studies, 1 prospective cohort study, 2 studies based on semi-structured interviews, 2 systematic literature reviews, and 1 article synthesizing five different retention measurement approaches.

The International Association of Physicians in AIDS Care panel reviewed a total of 325 studies to develop the recommendations for improving linkage to and retention in care. The specific 'recommendation 2' statement referenced three studies, including two retrospective cohort studies (one multisite) and a systematic review.

1a.10) Provide the estimates of benefit, and consistency across studies.

Studies consistently showed poor retention associated with greater risk of death. Studies also found that two approaches to measuring retention, constancy measures (e.g., based on number of visits per year) and missed visit metrics, are both valid, predictive of survival, and useful as clinical quality measures.

1a.11) Indicate what, if any, harms were identified in the study.

None identified.

1a.12) Identify any new studies conducted since the systematic review, and

indicate whether the new studies change the conclusions from the systematic review.

None identified.

Evidence

1a.13) If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.

N/A

1a.14) Briefly synthesize the evidence that supports the measure.

N/A

1a.15) Detail the process used to identify the evidence.

N/A

1a.16) Provide the citation(s) for the evidence.

N/A

Importance to Measure and Report: Gap in Care/Disparities (1b.01 - 1b.05)

1b.01) Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care and list the benefits or improvements in quality envisioned by use of this measure.

The HIV "continuum of care" is the process of HIV testing, linkage to HIV care, initiation of ART, adherence to treatment, retention in care, and virologic suppression (Gardner et al 2011). Poor retention in care is associated with lower rates of ART use (Giordano et al 2003), delayed viral suppression (Crawford et al 2014), and increased risk of mortality (Giordano et al 2007; Mugavero et al 2009). This measure will help providers direct their attention and quality improvement efforts towards improving retention in care.

CITATIONS:

Crawford, T. N., Sanderson, W. T., & Thornton, A. (2014). *Impact of Poor Retention in HIV Medical Care on Time to Viral Load Suppression. Journal of the International Association of Providers of AIDS Care (JIAPAC)*, 13(3), 242–249.
<https://doi.org/10.1177/2325957413491431>.

Gardner, E. M., McLees, M. P., Steiner, J. F., Del Rio, C., & Burman, W. J. (2011). *The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 52(6), 793–800.
<https://doi.org/10.1093/cid/ciq243>.

Giordano, T. P., Gifford, A. L., White, A. C., Suarez-Almazor, M. E., Rabeneck, L., Hartman, C., Backus, L. I., Mole, L. A., & Morgan, R. O. (2007). *Retention in care: A challenge to survival with HIV infection. Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 44(11), 1493–1499.
<https://doi.org/10.1086/516778>.

Giordano, T. P., White, A. C., Sajja, P., Graviss, E. A., Arduino, R. C., Adu-Oppong, A., Lahart, C. J., & Visnegarwala, F. (2003). *Factors associated with the use of highly active antiretroviral therapy in patients newly entering care in an urban clinic. Journal of Acquired Immune Deficiency Syndromes (JAIDS)*, 32(4), 399–405.
<https://doi.org/10.1097/00126334-200304010-00009>

Mugavero, M. J., Lin, H.-Y., Willig, J. H., Westfall, A. O., Ulett, K. B., Routman, J. S., Abrams, S., Raper, J. L., Saag, M. S., & Allison, J. J. (2009). *Missed visits and mortality among patients establishing initial outpatient HIV treatment. Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 48(2), 248–256. <https://doi.org/10.1086/595705>

1b.02) Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and 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.

We calculated the measure performance scores at the clinician level for the measure performance period from January 1, 2021, to December 31, 2021. The total sample included 188 clinicians and 3,136 unique patients, of which 48 clinicians had at least 11 patients—the minimum sample size requirement outlined in the CMS cell suppression policy. The data for these 48 clinicians included 3,065 unique patients. We provide the distribution of the measure performance scores for all clinicians and clinicians with at least 11 patients eligible for the denominator below. In the remainder of this document, we will focus on clinicians with at least 11 patients.

Table 1b.02. Distribution of the measure performance scores in the clinician samples

Sample	Clinicians	Patients	Mean	SD	Min	p10	p20	p30	p40	p50	p60	p70	p80	p90	Max	IQR
All Clinicians	188	3,136	89.9	18.5	0.0	66.7	81.6	88.9	95.2	100.0	100.0	100.0	100.0	100.0	100.0	14.3
Clinicians with 11+ patients	48	3,065	89.7	8.2	58.3	80.4	83.6	86.7	88.4	91.4	92.5	94.1	96.6	100.0	100.0	9.5

Notes: SD=Standard deviation, Min=minimum, Max=maximum, p=percentile, IQR=interquartile range. These data reflect Ryan White HIV/AIDS Program clients, and thus reflect a sample of patients with higher rates of annual retention in care than the national population of people with HIV (HRSA 2022, CDC 2020).

ALT-TEXT:

Table 1b.02 describes the distribution of the measure scores in the sample of all clinicians and clinicians with at least eleven patients. The table shows the number of clinicians and patients, and the mean, standard deviation, interquartile range and minimum and maximum scores on the measure, as well as the distribution of the measure scores by decile. The table shows that there’s an overall high performance on the measure but there is a substantial variation in the measure scores indicating the potential for further improvement. Overall, there was slightly more variation and the higher percentile scores in the sample of all clinicians than in the sample of clinicians with at least eleven patients.

CITATIONS:

Health Resources and Services Administration. Ryan White HIV/AIDS Program Annual Client-Level Data Report 2021. (2022). www.hab.hrsa.gov/data/data-reports.

Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2018. (2020). HIV Surveillance Supplemental Report; 25(2).

<https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-25-2.pdf>

1b.03) If no or limited performance data on the measure as specified is reported above, 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. Include citations.

Not applicable.

1b.04) 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.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. 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.

Data for this analysis are the same as the measure testing data (see section sp.27 for full details.) During testing, the measure performance was stratified to assess whether there were disparities in retention in care by patients’ age (< 50 years vs. >= 50 years), HIV transmission category (men who have sex with men [MSM], injection drug use [IDU], which includes both IDU and MSM and IDU, and other transmission), race (White vs. Black), and ethnicity (Hispanic or Latino vs. not Hispanic or Latino). Table 1b.04 summarizes the results of the analysis.



Table 1b.04. Measure performance rates by population groups (age, HIV transmission, race, and ethnicity)

Patient Group	Clinicians	Patients	Mean	Std dev	Min	10 %	20 %	30 %	40 %	50%	60%	70%	80%	90%	Max	IQR
AGE<50	48	1,521	87.4	10.9	55.6	75.0	78.2	80.2	83.8	88.7	90.9	92.6	100.0	100.0	100.0	21.3
AGE>=50	48	1,544	91.5	10.4	50.0	77.5	86.4	90.4	91.3	93.9	95.8	100.0	100.0	100.0	100.0	11.7
MSM	42	1,228	91.0	10.0	57.1	77.1	84.8	88.3	89.3	91.4	95.8	100.0	100.0	100.0	100.0	13.1
IDU	41	239	84.3	25.3	0.0	50.0	75.0	80.8	87.5	100.0	100.0	100.0	100.0	100.0	100.0	21.9
Other transmission	42	1,005	90.8	9.2	68.2	76.5	83.3	89.9	90.5	92.9	94.1	99.2	100.0	100.0	100.0	15.6
Black	48	1,399	86.7	17.1	0.0	76.3	80.2	83.4	86.4	90.5	92.9	96.1	100.0	100.0	100.0	18.4
White	48	1,542	90.6	14.3	25.0	78.3	86.0	90.5	91.7	93.9	100.0	100.0	100.0	100.0	100.0	12.6
Not Hispanic or Latino	48	2,622	89.4	9.0	58.3	78.7	81.9	84.3	87.0	91.0	92.4	96.3	99.1	100.0	100.0	14.2
Hispanic or Latino	37	436	93.2	10.7	50.0	79.4	86.3	92.0	94.1	100.0	100.0	100.0	100.0	100.0	100.0	11.1

Notes: Results are for clinicians with ≥11 patients eligible for the denominator.
SD=standard deviation, min=minimum, max=maximum, IQR=interquartile range

ALT-TEXT:

Table 1b.04 describes the measure performance rate age, virus transmission model, race, and ethnicity. The table shows the number of clinicians and patients, and the mean, standard deviation, interquartile range, and minimum and maximum scores on the measure, as well as the distribution of the measure scores by decile. The measure scores are shown separately for 1) patients less than 50 years of age, patients fifty years of age and older, 2) patients with MSM, IDU and other (i.e.: non-MSM/IDU) virus transmission model, 3) White and Black patients, and 4) Hispanic or Latino and not Hispanic or Latino patients. The mean measure scores are higher for patients fifty years of age and older, patients with non-IDU or MSM transmission methods, White patients and Hispanic or Latino patients.

The measure rate differences by group are in the expected direction for race, age, and HIV transmission group but are not statistically significant (HRSA 2022), likely because there are relatively high rates of retention overall among Ryan White recipients and somewhat mixed evidence of disparities by race, age, and HIV transmission group in more recent studies (Anderson 2020, Kay and Westfall 2020). Across all clinicians, the mean clinician-level performance rates were similar for all strata, though we observed more variability in the clinician-level measure rates for the MSM/IDU group of patients (mean=84.3, SD=25.3). Larger variation in the measure rates for the MSM/IDU patients can be attributed to the relatively small sample for that category. However, the differences in the measure scores by the HIV transmission mode were not significant ($p=0.96$). We also did not find statistically significant differences in the measure scores by race ($p=0.23$) or ethnicity ($p=0.08$). The measure rates were lower for the younger (less than 50-year-old) patients, which reflects the lower rates of retention in care among younger patients based on national surveillance data (HRSA 2022). However, the differences in the measure rates by age did not reach the $p\leq 0.05$ threshold of statistical significance ($p=0.07$).

CITATIONS:

Anderson, A.N., Higgins, C.M., Haardörfer, R. et al. Disparities in Retention in Care Among Adults Living with HIV/AIDS: A Systematic Review. AIDS Behav 24, 985–997 (2020). <https://doi.org/10.1007/s10461-019-02679-2>

Health Resources and Services Administration. Ryan White HIV/AIDS Program Annual Client-Level Data Report 2021. (2022). <https://ryanwhite.hrsa.gov/data/reports>

Emma Sophia Kay & Andrew O. Westfall (2020) Ryan White HIV/AIDS program recipients more likely than non-recipients to be retained in care using six different retention measures, AIDS Care, 32:1, 89-92, DOI: [10.1080/09540121.2019.1623375](https://doi.org/10.1080/09540121.2019.1623375)

1b.05) If no or limited data on disparities from the measure as specified is reported above, 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 above.

Not applicable.

Scientific Acceptability: Maintenance (2ma.01 - 2ma.04)

2ma.01) Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

Yes

No

2ma.02) Indicate whether additional empirical validity testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

Yes

No

2ma.03) For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?

- Yes
- No

2ma.04) For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

- Yes - Additional risk adjustment analysis is included
- No additional risk adjustment analysis included

Scientific Acceptability: Reliability - Testing (2a.01 - 2a.12)

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 fields in the Scientific Acceptability sections of the Measure Submission Form.

- 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 Battelle staff at PQMsupport@battelle.org about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact Battelle staff at PQMsupport@battelle.org with any questions.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the [2021 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 the evaluation criteria for testing.

2a. Reliability testing 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 instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing 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 instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) 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).

2b3. 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 social risk 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.

2b4. 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.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. 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 non-responders) 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)

Definitions

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).

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 measure 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. The degree of consensus and any areas of disagreement must be provided/discussed.

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.

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

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

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.

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

2a.01) Select only the data sources for which the measure is tested.

- Assessment Data
- Claims
- Electronic Health Data
- Electronic Health Records
- Instrument-Based Data
- Management Data
- Other (please specify here:)
- Paper Medical Records
- Registry Data

2a.02) 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 submission uses the structured EHR patient-level data from seven participating sites. Please refer to section sp.27 for more details about the sample.

2a.03) Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

01-01-2021 – 12-31-2021

2a.04) Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

- Accountable Care Organization
- Clinician: Group/Practice
- Clinician: Individual

- Facility
- Health Plan
- Integrated Delivery System
- Other (specify)
- Population: Community, County or City
- Population: Regional and State

2a.05) List the measured entities 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.

Seven test sites that are Ryan White HIV/AIDS Program recipients representing three regions (Northeast, South, and Midwest) provided the data for this measure. Of these seven sites, four test sites were publicly funded community health centers, two sites were hospital-based clinics, and one site represented a community-based service organization. The sites varied in EHR systems (eClinical Works, EPIC/OCHIN EPIC, NextGen, Athena Health). At these 7 test sites, a total of 48 clinicians were included in testing. These 48 clinicians had a total of 3,065 patients included in the measure denominator. Table 2a.05 breaks down the characteristics of the participating sites included in the beta testing of the measure.

Table 2a.05. Test site characteristics

Site	Provider type	Region	EHR	# of clinicians with 11+ patients	# of patients
Site 1	Publicly funded community health center	NE	eClinical Works	5	137
Site 2	Publicly funded community health center	NE	eClinical Works	3	159
Site 3	Hospital or university-based clinic	NE	EPIC	22	607
Site 4	Publicly funded community health center	SO	OCHIN EPIC	6	558
Site 5	Other community-based service organization	NE	NextGen	3	62
Site 6	Publicly funded community health center	MW	Athena Health	3	492
Site 7	Hospital or university-based clinic	MW	eClinical Works	6	1050

Notes: NE=Northeast, SO=South, MW=Midwest

ALT-TEXT:

Table 2a.05 provides characteristics of the seven test sites, including provider type,

provider region, provider electronic health record system, as well as the number of clinicians (for clinicians with at least 11 patients) and patients in the sample the measure developer received from each provider.

2a.06) Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

The full analytic sample included 3,136 patients attributed to 188 clinicians within the measurement period from 7 sites that are Ryan White HIV/AIDS Program participants. The measure is specified to require a minimum denominator of 11 patients during the measurement period, to follow the CMS cell size suppression policy. The policy sets minimum thresholds for the display of CMS data which states no cell can be reported that allows a value of 1 to 10 reported. The restricted sample includes 48 clinicians (25.5% of the initial number of clinicians) and 3,065 patients (97.7% of the initial number of patients). When limited clinicians with 11 or more patients eligible for the denominator during the measurement period, the average (mean) clinician has a measure rate of 89.7%.

Out of the patients attributed to a clinician with at least 11 patients in the denominator:

- 3% (93) patients were under the age of 25 and 97% (3,2972) patients were over the age of 25
- 50% (1,521) patients were under the age of 50, and 50% (1,544) patients were over the age of 50.
- Broken out by HIV transmission group, 40% (1,228) of patients' transmission group was men who have sex with men (MSM), 8% (239) of patients' transmission group was injection drug use (IDU; note that IDU included patients who had both MSM and IDU listed as their HIV transmission group), and 33% (1,005) transmission group was other, while 19% (593) of patients were missing information on HIV transmission group.
- 25% (771) patients were cisgender women, 74% (2,279) patients were cisgender men, and 0.5% (15) patients were transgender women.

2a.07) 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.

The same data are used for all clinician-level testing (reliability, concurrent validity, known group validity, and meaningful difference in performance), as described below.

Reliability: To assess reliability, we used the EHR data from each of seven sites

covering the period between January 1, 2021, and December 31, 2021. For the reliability analysis (and all other clinician-level testing) we restricted the sample to clinicians who saw at least 11 patients during the measurement period.

Data Element Validity: To assess data element validity, we randomly selected a subset of 20 patient encounters (from the full EHR extract) in each of the seven sites, for a total of 140 encounters. For selected cases, site personnel manually abstracted data elements necessary for the measure calculation from each site's EHR. We then compared the manually abstracted and electronically extracted data to assess data element validity via agreement between the gold-standard source (manual abstraction) and the EHR extract.

Construct validity: To assess validity of the measure using known-group validity method we stratified the sample by age (patients <50 years old vs. patients 50 years old or older) and HIV transmission group (IDU vs. non-IDU). For this analysis we used the EHR data from each of seven sites covering the period between January 1, 2021, and December 31, 2021, and we restricted the sample to clinicians who saw at least 11 patients during the measurement period.

Face validity: We solicited feedback on the measure's face validity from 7 clinicians via a semi-structured interview. We also conducted a formal poll during a meeting of the Technical Expert Panel (TEP) that was convened during the development of the HIV Annual Retention in Care measure. The TEP was comprised of clinicians, patient representatives, and other experts in EHR systems and HIV care.

Exclusions: Not applicable; this measure does not have exclusions.

Risk adjustment: Not applicable; this measure is not risk adjusted.

Meaningful difference in performance: To assess whether meaningful differences in the measure performance we restricted the sample to clinicians who saw at least 11 patients during the measurement period, using EHR data from each of seven sites covering the period between January 1, 2021, and December 31, 2021.

2a.08) List the social risk factors that were available and analyzed.

For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter “see validity testing section of data elements”; and enter “N/A” for 2a.11 and 2a.12.

Not applicable. This is a process measure.

2a.09) Select the level of reliability testing conducted.

Choose one or both levels.

- Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)
- Accountable Entity Level (e.g., signal-to-noise analysis)

2a.10) For each level of reliability testing 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.

We tested reliability of the measure at the clinician level using three methods: signal-to-noise reliability, which tests the precision of the measure rates at the clinician level, and split-half and test-retest methods, which test the stability of the measure rates across clinicians.

The signal-to-noise method summarizes the proportion of the total variation in the clinician scores that is attributable to real underlying differences between clinicians (signal), in relation to random variation within each clinician (noise). Noise can be introduced by patient-level variability, which might include unmeasured patient characteristics, or by the lack of precision in the measure estimates because of a lack of sufficient patient sample size within clinicians (Deutsch et al 2012). The beta-binomial model is an appropriate framework for estimating reliability for the measure (Adams 2009). Reliability is calculated as the ratio of the variance between clinicians and the total variance (that is, the sum of the between-clinician and within-clinician variances) of the measure rates.

The resulting reliability statistic ranges from 0 to 1. If reliability is 0, there is no variation on the measure across clinicians, and all observed variation is because of random variation within a clinician. In this case, the lack of reliability suggests that the measure is not useful for distinguishing between clinicians with respect to that outcome. Conversely, if reliability is 1, all provider scores are free of random variation, and all variation represents real differences between clinicians in the measure result.

After we computed the signal and noise variance for providers in the sample, we determined the minimum denominator size necessary to reach the reliability of 0.7, which is commonly considered the threshold for acceptable reliability.

We used split-half and test-retest approaches to examine stability of the measure scores within the same clinicians. The split-half method involves comparing the measure scores for two independent samples of patients within the same provider. For this method, we randomly split the sample of patients within each clinician into two

mutually exclusive samples with equal or nearly equal size, resulting in two samples that cover the same one-year period but with case volume the size of a measure that would be calculated with six months of data. **Thus, each clinician appeared in the sample twice, but with an entirely different set of patients.** Then, we estimated Spearman- and intraclass correlations between the measure rates within two samples. We also computed the Spearman-Brown correction to account for the attenuation of the Spearman correlation due to dividing the original sample of patients in two halves.

Since split-half method can, under some conditions, over- or underestimate reliability because of capitalization on chance, we also assessed stability of the measure scores using the test-retest method, which involves comparing the measure scores for the same clinicians computed in different samples of patients either within the same measurement period or two adjacent measurement periods. Since we only had one year of data, we opted to use bootstrap resampling to generate independent samples of patients within the same clinicians. The bootstrap method avoids biased sampling, maintains the original sample size, and allows estimation of confidence intervals for the reliability estimates. We drew 2,000 independent samples with replacement (stratified by the provider), maintaining the same number of beneficiaries for each provider as in the original sample, and grouped the samples into 1,000 pairs. These random samples from a given clinician are assumed to reflect an independent set of re-measurement of the HIV Annual Retention in Care rates for a clinician. Then, we estimated Spearman- and intraclass correlations between the measure rates within each pair and computed the mean and the 95percent confidence interval of the distribution of correlations from the 1,000 paired samples.

Spearman correlation captures the association between the ranks of clinicians in different realizations of the bootstrap samples. The intraclass correlation captures the degree of correlation and agreement between measurements and is represented as a ratio of the variance in the measure counts between providers over the sum of the variances between and within providers. Hence, the smaller the disagreement between the measure counts for each clinician in different samples, the larger the intraclass correlation coefficient. Correlation values range from 0 to 1; a value of 1 indicates perfect reliability, and a value of 0 means the measure is perfectly unreliable.

Following CMS's cell size suppression policy for reporting, all clinicians with fewer than 11 patients in the measurement period were excluded from calculations.

CITATIONS:

Adams, J. L. (2009). *The Reliability of Provider Profiling: A Tutorial*. https://www.rand.org/pubs/technical_reports/TR653.html

Deutsch, A., Smith, L., Gage, B., Kelleher, C., & Garfinkel, D. (2012). *Patient-Reported Outcomes in Performance Measurement*. https://www.qualityforum.org/Projects/n-r/Patient-Reported_Outcomes/Commissioned_Paper_2.aspx

2a.11) For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, Measure Evaluation Criteria).

Table 2a.11-A summarizes the mean and range of the signal-to-noise reliability statistics for the measure, which was calculated separately by facility. The mean signal-to-noise reliability across all 48 clinicians with at least 11 patients eligible for the denominator exceeds the 0.70 threshold for acceptable reliability. The 5th percentile for the measure reliability was 0.77, and the 75th percentile was 0.99. These results suggest that the measure is highly reliable.

Table 2a.11-A Signal to Noise Reliability of Clinician-level Measure Scores

Sample	Clinicians	Mean	min	p5	p10	p25	p50	p75	p90	p95	max	SD
Clinicians >= 11 patients	48	0.94	0.64	0.77	0.82	0.92	0.97	0.99	1.00	1.00	1.00	0.08

Notes: Results are provided for clinicians with ≥11 patients eligible for the denominator (n = 48). Min=minimum, p=percentile, max=maximum, SD=standard deviation.

ALT-TEXT:

Table 2a.11-A provides results of the signal-to-noise reliability testing for clinicians with at least 11 patients eligible for the denominator. The table shows the number of clinicians, and the mean, standard deviation, minimum and maximum scores on the measure, as well as the measure scores for the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles. Mean reliability in a sample was very high and measure scores for most clinicians in a sample were highly reliable.

Table 2a.11-B summarizes the Spearman rank-order correlation, Spearman-Brown correlation, and intra-class correlation for the split-half reliability statistics for the HIV Annual Retention in Care measure. All statistics exceeded the 0.9 threshold indicating very high stability of the measure scores across independent samples of patients.

Table 2a.11-B Split-Half Reliability of Clinician-level Measure Score

Sample	Spearman rank-order correlation	Spearman-Brown correction for the Spearman correlation	Intra-class correlation
Clinicians >= 11 patients	0.975	0.987	0.946

Notes: Results are provided for clinicians with ≥ 11 patients eligible for the denominator (n = 48)

Table 2a.11-C summarizes the mean and range of the Spearman rank-order correlation, Spearman-Brown correlation, and intra-class correlation for the test-retest reliability statistics for the HIV Annual Retention in Care measure, which was calculated using the bootstrap method in 1,000 pairs of bootstrap samples. In this analysis, the mean reliability statistics (e.g.: mean intraclass correlations across 1,000 samples) exceeded the 0.9 threshold indicating very high stability of the measure scores across 1,000 pairs of patient samples.

ALT-TEXT:

Table 2a-11B provides the results of the split-half reliability for clinicians with at least 11 patients eligible for the denominator. Spearman correlation, Spearman correlation corrected for sample attenuation using Spearman-Brown correction, and intra-class correlations exceeded 0.9 indicating high split-half reliability of the measure.

Table 2a.11-C Test-Retest Reliability of Clinician-level Measure Scores via the Bootstrap Resampling Method

Sample	Statistical Method	Mean	Min	p5	p10	p25	p50	p75	p90	p95	Max
Clinicians ≥ 11 patients	Spearman rank-order correlation	0.983	0.919	0.971	0.975	0.980	0.984	0.988	0.990	0.992	0.995
Clinicians ≥ 11 patients	Spearman-Brown correction	0.991	0.958	0.985	0.987	0.990	0.992	0.994	0.995	0.996	0.998
Clinicians ≥ 11 patients	Intra-class correlation	0.972	0.904	0.948	0.955	0.965	0.974	0.981	0.985	0.987	0.992

Notes: Results are provided for clinicians with ≥ 11 patients eligible for the denominator (n = 48)

ALT-TEXT:

Table 2a.11-C shows the results for test-retest reliability of the measure scores estimated using bootstrap resampling for clinicians with at least 11 patients eligible for the denominator. The table shows mean, minimum, maximum and the percentile distribution of the Spearman correlations, Spearman correlations corrected for sample attenuation using Spearman-Brown correction, and intra-class correlations obtained in 1,000 pairs of samples generated using bootstrap. The mean reliability exceeded 0.9 for all three methods indicating high test-retest reliability of the measure.

Our findings regarding the minimum sample size for the measure indicate the median sample size per clinician necessary to reach signal-to-noise reliability of 0.7 in our

sample was 7 patients. With the sample size of 11 patients, 95 percent of clinicians would reach the 0.7 threshold for signal-to-noise reliability. Therefore, assuming our findings are generalizable to the universe of reporting clinicians, using the CMS recommendations for the minimum sample size for reporting purposes would allow clinicians to achieve the reliability of at least 0.7 for 95 percent of clinicians in our sample. Our findings thus indicate that following CMS recommendations for the minimum sample size for reporting purposes (11+ patients) would allow us to produce statistically valid comparisons between clinicians.

2a.12) Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

The HIV Annual Retention in Care measure demonstrates high reliability in all three methods of testing indicating that the measure could be useful to distinguish a clinician's performance from the sample mean and between any clinician pair. The results also indicate that the measure has good stability.

Although there is not a clear cut-off for the minimum signal-to-noise reliability level, reliability of 0.4 is often considered to be the lower limit of moderate reliability sufficient for public reporting (Schone, Hubbard and Jones, 2011), reliability above 0.7 is considered sufficient to see differences between physicians and the mean (Adams, 2009), and reliability above 0.9 is considered sufficient to see differences between any physician pair (National Quality Forum, 2013). According to our calculations, not only is the measure's average reliability high, but most individual clinicians also have highly reliable scores.

The ICC captures the effect of the clinician on the patients' outcomes and could be interpreted as the correlation in the outcome between two individuals randomly selected from the same clinician (Austin and Merlo, 2017). There are no standard values for acceptable reliability using ICC. A low ICC could not only reflect the low degree of agreement but also relate to the small number of subjects. Following Porteny and Watkins, we rely on the following interpretation: ICC values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability. The ICCs exceeding 0.9 obtained from the split-half and test-retest testing methods indicate excellent reliability of the measures.

Finally, according to Cohen's effect-size criteria, the Spearman correlations above 0.8 indicate a large effect size, thus also supporting the claim that the measure is highly reliable.

CITATIONS:

Adams JL. The Reliability of Provider Profiling: A Tutorial. Santa Monica, CA: RAND

Corporation; 2009. http://www.rand.org/pubs/technical_reports/TR653.html.
doi:10.7249/TR653

Austin, P. C., & Merlo, J. (2017). Intermediate and advanced topics in multilevel logistic regression analysis. *Statistics in Medicine*, 36(20), 3257–3277.
<https://doi.org/10.1002/sim.7336>

Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155.

Deutsch A, Smith L, Gage B, Kelleher C, Garfinkel D. (2012) Patient-reported outcomes in performance measurement. https://www.qualityforum.org/Projects/n-r/Patient-Reported_Outcomes/Commissioned_Paper_2.aspx

National Quality Forum. *Guidance for Measure Testing and Evaluating Scientific Acceptability of Measure Properties* (2011).
<https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=70943>.

Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. New Jersey: Prentice Hall; 2000

Schone E, Hubbard M, Jones D. (2011). Reporting period and reliability of AHRQ, CMS 30-day and HAC quality measures. Memorandum submitted to the Centers for Medicare and Medicaid Services.
<https://EconPapers.repec.org/RePEc:mpr:mprres:cab712bf5e324d0db15eca9c404f3eb2>.

Scientific Acceptability: Validity - Testing (2b.01 - 2b.04)**2b.01) Select the level of validity testing that was conducted.**

- Patient or Encounter-Level (data element validity must address ALL critical data elements)
- Accountable Entity Level (e.g., hospitals, clinicians)
- Empirical validity testing of the measure score
- 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)

2b.02) 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.

We tested the data element validity, face validity (qualitative assessment of validity based on the experts' feedback), and measure score validity (construct validity) for all three eCQM.

Data element validity. This form of validity testing assesses whether the data elements, as obtained from the structured, extractable fields in the EHR, accurately reflect the care the patients received by comparing agreement between the gold-standard assessment of care (manual chart review) with the EHR extract for a randomly selected set of patients:

1. EHR data extracted from the structured fields used by the eCQMs
2. Manually abstracted data from the entire medical record, including free-text note fields and scanned documents

We requested that the test sites pull the structured patient data from their EHRs through automated extraction algorithms and send us the data. To support this request, we provided sites with data dictionaries containing all of the data elements associated with the three measures, as well as the corresponding value sets. We then selected a random sample of medical record numbers from the EHR extract from each site (20 records per site for a total of 140 records across 7 sites) and ask the site staff to abstract the same data elements through a manual review of the patients' medical records. At all phases of the EHR extract and manual abstraction process we met with sites as needed to answer questions about the process.

We calculated the raw agreement (percentage agreement) and the chance-corrected agreement (Gwet's AC1) between the two data sources. The interpretation of the AC1 statistic is the same as that of Cohen's Kappa, but AC1 is a more robust measure of interrater reliability. Kappa is sensitive to classification probabilities which in some cases lead to the low chance-corrected agreement despite the high observed agreement (the so-called Kappa paradox). This situation does not occur when using AC1 (Quarfoot and Levine 2016). Higher values for agreement statistics demonstrate that the structured EHR data used to calculate the measure have accuracy similar to looking at the medical record overall, including clinical notes, documents, and other fields that convey information about the patient but cannot be used to calculate eQMs. When the two measurements agree perfectly, the value of the agreement will be 1.0.

Face validity: We conducted clinician interviews with seven clinicians from the seven test sites. We developed an interview guide to solicit clinician perspectives on the utility and face validity of the measure. Specifically, we asked whether they thought measure scores could be used to accurately distinguish quality among providers. The evaluation of face validity was conducted through a semi-structured interview process. We also conducted a formal poll during a meeting of the Technical Expert Panel (TEP) that was convened during the development of the HIV Annual Retention in Care measure. The TEP was comprised of clinicians, patient representatives, and other experts in EHR systems and HIV care.

Construct validity. We assessed the differences in the measure rates by subgroups shown in the literature to have differences in rates of retention in HIV care: transmission category (injection drug use versus non-injection drug use) and age (younger than age 50 versus 50 or older). This approach, known-group validity, is a hypothesis-based testing that leverages hypotheses based on known differences in care to see if the same differences are reflected in the measure rates, thus providing evidence of the measure's validity. For each characteristic, we stratified the sample, calculate the measure rates, and compute the effect size using Cohen's *d* statistic. A higher absolute value of Cohen's *d* indicates a higher standardized difference between the two groups. NQF does not set specific thresholds for known-group validity; rather, the committee might consider collective evidence from all validity tests to adjudicate the measure.

CITATIONS:

Quarfoot, D., & Levine, R. A. (2016). *How Robust Are Multirater Interrater Reliability Indices to Changes in Frequency Distribution? The American Statistician*, 70(4), 373–384. <https://doi.org/10.1080/00031305.2016.1141708>

2b.03) Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

Data element validity. We found that the percent agreement between the EHR data

and manual abstraction ranged from 75% to 100%, corresponding to a Gwet's AC1 ranging from 0.68 to 1.00.

Table 2b.03-A. Agreement Between Medical Records and Manual Abstraction

Data Element	Percent Agreement	Gwet's AC1
HIV diagnosis date	75.0	0.68
Encounter 1 date	99.3	0.99
Encounter 1 type	99.3	0.99
Encounter 2 date	99.3	0.99
Encounter 1 type	99.3	0.99
Viral load 1 date	98.6	0.98
Viral load 1 value	98.6	0.98
Viral load 2 date	98.6	0.97
Viral load 2 value	100.0	1.00
Average across all elements	96.4	0.95

ALT-TEXT:

Table 2b.03-A shows the observed agreement expressed as percent agreement and the chance-corrected agreement computed using Gwet's AC1 statistic between the medical records and manual abstraction across 140 records collected from 7 sites. Results indicate very high observed and change-corrected (above 0.95) reliability for all data elements except for the HIV diagnosis date, for which the observed reliability was 0.75 and chance-corrected reliability was 0.68.

Face validity. We found that four out of seven (57%) clinicians agreed that the measure can distinguish between good and poor quality of care and that five out of seven (71%) agreed with the measure specifications. We also found that 88% of the Technical Expert Panel agreed that the HIV Annual Retention in Care measure was important and related to quality of care.

Construct validity. Among patients attributed to clinicians with at least 11 patients, the mean rate of retention in care was 87.4% for those under 50 as compared to 91.5% for those 50 years or older with an effect size of -0.381 using Cohen's D, and the mean rate of retention in care was 83.5% for those whose HIV transmission group was IDU as compared to 90.7% for those whose HIV transmission group was not IDU, with an effect size of 0.382 using Cohen's D.

Table 2b.03-C. Known-group validity results for clinicians ≥ 11 patients eligible for denominator

Patient sub-group	Mean Retention in Care	S.D.	Cohen's D	t-test p-value
Age < 50 years	87.4	10.9	0.381 (results for comparison of age sub-groups)	0.065 (results for comparison of age sub-groups)
Age ≥ 50 years	91.5	10.4	*	*
Non-IDU HIV transmission category	90.7	7.9	0.382 (results for comparison of HIV transmission category subgroups)	0.100 (results for comparison of HIV transmission category subgroups)
IDU HIV transmission category	83.5	25.7	*	*

Notes: * = Cell intentionally left empty; S.D. = standard deviation

ALT-TEXT:

Table 2b.03-C shows results for the known-group validity testing for clinicians with at least 11 patients eligible for the denominator. The table shows the mean retention in care scores and standard deviations, as well as Cohen's D statistics and p-values for the t-tests by groups. The results are provided separately for patients by age groups (less than 50 years of age, and 50 years and older), and patients with non-IDU and IDU HIV transmission category. The mean measure scores were higher for the patients 50 years and older and patients with non-IDU HIV transmission category.

2b.04) Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)

Face validity. Most respondents agreed with the measure specifications (5 out of 7) and the majority of respondents strongly agreed or agreed that the measure score is an accurate reflection of quality and that the measure score can be used to distinguish between good and poor quality of care (4 out of 7). Among the technical expert panel, the vast majority (88 percent) agreed that the measure was important and related to quality of care. These results demonstrate acceptable face validity of the measure.

Data element validity. The AC1 values calculated through data element validity testing suggest high levels of agreement between the data extract generated from the EHR systems and the manually abstracted data. We observed the 96.4 percent average agreement across all data elements or higher for all data elements. The average

chance-corrected agreement captured by the AC1 statistic was 95.2 percent. These statistics indicate very high data element validity. Observed agreement was very high (above 98 percent) for all data elements with the exception for the HIV diagnosis date, for which the observed agreement was 75 percent. As a sensitivity test, we compared the HIV diagnosis dates in both data sources (i.e., EHR and chart data). We found that in all cases, observed differences between the HIV diagnosis dates did not affect whether or not the patient would have been included in the denominator because all the dates occurred prior to the measurement period.

Known-group validity

We found that measure rates were about 4 percentage points higher in patients who are age 50 and older as compared to those who are below 50 years old and that rates were about 7 percentage points lower among those whose transmission group was IDU as compared to non-IDU. These results align with prior findings based on a systematic review of retention in care literature and provide evidence of measure validity (Anderson et al 2020).

CITATIONS:

Anderson, A. N., Higgins, C. M., Haardörfer, R., Holstad, M. M., Nguyen, M. L. T., & Waldrop-Valverde, D. (2020). Disparities in Retention in Care Among Adults Living with HIV/AIDS: A Systematic Review. AIDS and behavior, 24(4), 985–997.
<https://doi.org/10.1007/s10461-019-02679-2>

Scientific Acceptability: Validity - Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) (2b.05 - 2b.14)

2b.05) 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 in Importance to Measure and Report: Gap in Care/Disparities.

To examine differences in performance, we calculated measure performance rates across 47 clinicians with at least 11 patients eligible for the denominator in the performance period. We excluded clinicians with fewer than 11 patients eligible for the denominator following the CMS's cell suppression rule, and because the estimates for clinicians with fewer cases tend to be less reliable. Then, we computed a 95 percent confidence interval (95% CI) around each clinician's measure score and compared the 95% CI to the mean measure rate in our sample. If the confidence intervals did not overlap with the mean measure rate in a sample, the clinician's performance was identified as significantly better or worse from the mean.

We also calculated the distributions of the measure rates to determine if the measure was "topped out." For the measure to be topped out, two conditions had to be met (Analysis of Topped-Out Measures 2014). First, the 75th performance percentile must be statistically indistinguishable (within two standard errors) from the 90th percentile. Second, the truncated coefficient of variation (TCV) (calculated by first removing the lower and upper 5th percentiles and then dividing the standard deviation by the mean of this truncated distribution) must be less than or equal to 0.10.

CITATIONS:

"Analysis of Topped-Out Measures Finalized for the PY 2016 ESRD QIP." Updated June 19, 2014. Available at <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/Downloads/AnalysisofTopped-OutMeasuresFinalizedforthePY2016ESRDQIP.pdf>. Accessed on December 8, 2022.

2b.06) Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include 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.

Based on the sample of 48 clinicians with at least 11 patients eligible for denominator, the measure rates in our sample ranged from 58.3% to 100.0% (with a median of 91.4%). Thus, there is variation in measure scores across facilities.

Table 2b.06-A. Performance Distribution of the Annual Retention in Care measure rates

Sample	Clinicians	Patients	Mean	SD	Min	p10	p25	p50	p75	p90	Max	IQR
Clinicians 11+ patients	48	3,065	89.7	8.2	58.3	80.4	84.9	91.4	94.5	100.0	100.0	9.5

Notes: SD=standard deviation, min=minimum, max=maximum, p=percentile, IQR=interquartile range

ALT-TEXT:

Table 2b.06-A shows the performance distribution of the measure scores for clinicians with at least 11 patients eligible for the denominator. The table shows the number of clinicians and patients for the sample of clinicians with at least 11 patients, as well as the mean, standard deviation, median, inter-quartile range, and percentile distribution of the measure scores in the sample.

Of the 48 clinicians, 11 (22.9% of all clinicians in the sample) performed significantly better and 5 clinicians (10.4%) performed worse than the sample average, which is conceptually equivalent to an “average-performing clinician” in a sample. Distribution of the performance categories shown in Table 2b.06-B suggests that improvement in the measure scores is possible for more than three quarters of clinicians whose performance scores were either no different from the sample average (66.7%) or worse than the sample average (10.4%).

Table 2b.06-B. Performance distribution of measure rates relative to the sample average for clinicians with at least 11 patients eligible for the denominator

Performance group	N and % of facilities	Mean performance rate
Better than the national rate	11 (22.9%)	99.1%
No different than the national rate	32 (66.7%)	88.5%
Worse than the national rate	5 (10.4%)	76.0%
All Clinicians	48 (100.0%)	89.7%

ALT-TEXT:

Table 2b.06-B shows performance distribution of the HIV Annual Retention in Care measure scores relative to the sample average for clinicians with at least 11 patients

eligible for the denominator. Out of 48 clinicians in the sample, 32 clinicians had measure scores that were not significantly different from the sample average and 11 and 5 clinicians had measure scores that were, respectively, better and worse than the sample average.

The results of the topped-out analysis indicate that the measure’s 75th percentile is not within two standard deviations of the 90th percentile, thus it does not meet criterion 1 for being topped out. However, the measure has a truncated coefficient of variation (TCV) equal to 0.07 (therefore <0.10), which meets the second criterion. Because the measure does not meet both criteria, it is not considered topped out.

Table 2b.06-C. Topped out analysis of the HIV Annual Retention in Care measure rates.

Measure	75 th pctl.	90 th pctl.	90 th – 75 th pctl.	2x S.D. of 90 th pctl.	Criterion 1 met?	TCV	Criterion 2 met?
HIV Annual Retention in Care	0.94	1.00	0.06	0.01	No	0.07	Yes

Notes: pctl= percentile, S.D= standard deviation, TCV= truncated coefficient of variation.

Table 2b.06-C shows results of the topped-out analysis of the measure scores. For the measure to be topped out, two criteria must be met. First, the truncated coefficient of variation must be less than 0.10, and second, the 90th percentile on the measure score distribution must be indistinguishable from the 75th percentile. The table shows the results for both statistical analyses.

2b.07) Provide 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.

In other words, what do the results mean in terms of statistical and meaningful differences?

We observed substantial variability in measure rates across clinicians. The measure was also able to differentiate between clinicians with better and worse than average performance. The results indicate ample room for improvement and meaningful differences in the quality of care between the highest and lowest performing clinicians. The measure is also not considered topped out based on testing. Further, these data reflect Ryan White HIV/AIDS Program patients, and thus reflect a sample of patients with higher rates of retention in care than the national population of people with HIV (HRSA 2022, CDC 2020). It is likely that when this measure is applied to a broader

population of patients with HIV, the performance scores are unlikely to be as high.

CITATIONS:

Health Resources and Services Administration. Ryan White HIV/AIDS Program Annual Client-Level Data Report 2021. (2022). <https://ryanwhite.hrsa.gov/data/reports>.

Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2018. (2020). HIV Surveillance Supplemental Report; 25(2). <https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-25-2.pdf>.

2b.08) Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

Data elements required to calculate the performance rate are ones in which absence of data in a data field reflects the absence of an eligible encounter or laboratory test. For example, if a lab visit field had a missing value for a given patient, we interpret this to mean that a patient did not have an eligible lab visit, rather than the information for that visit was missing.

2b.09) Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

Because the measure is based on the presence of encounters, viral load tests, and HIV diagnosis dates, there were no other fields that had missingness to assess and the lack of observations in these fields indicated that no encounter, test, or diagnosis was present.

2b.10) Provide 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 non-responders), and how the specified handling of missing data minimizes bias.

In other words, 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 was conducted, justify the selected approach for missing data.

Note: This item is directed to measures that are risk-adjusted (with or without social risk 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 eCQMs). 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 social risk 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.

Our analysis indicates that missing data are not a threat to validity for the measure among the data elements in the fields which are required for the measure calculations.

2b.11) Indicate whether there is more than one set of specifications for this measure.

- Yes, there is more than one set of specifications for this measure
 No, there is only one set of specifications for this measure

2b.12) 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. Indicate what statistical analysis was used.

Not applicable.

2b.13) Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

Not applicable.

2b.14) Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

Not applicable.

**Scientific Acceptability: Validity - Other Threats to Validity
(Exclusions, Risk Adjustment) (2b.15 - 2b.32)****2b.15) Indicate whether the measure uses exclusions.**

- N/A or no exclusions
- Yes, the measure uses exclusions.

2b.16) Describe the method of testing exclusions and what was tested.

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.

2b.17) Provide 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.

2b.18) Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, 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.

2b.19) Check all methods used to address risk factors.

- Statistical risk model with risk factors (specify number of risk factors)
- Stratification by risk category (specify number of categories)
- Other (please specify here:)
- No risk adjustment or stratification

2b.20) If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

Not applicable.

2b.21) If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

HIV Annual Retention in Care is a process measure and thus should not be risk adjusted.

2b.22) Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

- Published literature
- Internal data analysis
- Other (please specify here: Not applicable.)

2b.23) Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of $p < 0.10$ or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

Not applicable.

2b.24) Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

Not applicable.

2b.25) Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

Not applicable.

2b.26) 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 control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter “N/A” for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

Not applicable.

2b.27) Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

Not applicable.

2b.28) Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

Not applicable.

2b.29) Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

Not applicable.

2b.30) Provide the results of the risk stratification analysis.

Not applicable.

2b.31) Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

Not applicable.

2b.32) Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

Not applicable.

Feasibility (3.01 - 3.07)**3.01) Check all methods below that are used to generate the data elements needed to compute the measure score.**

- 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-10 codes on claims)
- Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
- Other (Please describe)

3.02) Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields. ALL data elements are in defined fields in electronic health records (EHRs)

- ALL data elements are in defined fields in electronic claims
- ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)
- ALL data elements are in defined fields in a combination of electronic sources
- Some data elements are in defined fields in electronic sources
- No data elements are in defined fields in electronic sources
- Patient/family reported information (may be electronic or paper)

3.03) 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 data elements not from electronic sources.

Not applicable.

3.04) Describe any efforts to develop an eCQM.

We have developed this measure as an eCQM, including developing the specification in the current standard, the Quality Data Model (QDM), and completing Bonnie testing with 100% passing and coverage.

3.05) Complete and attach the eCQM-Feasibility-Scorecard.xls file.**3.06) 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.

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

With one exception, all data elements required for this measure were consistently available and captured accurately across all seven test sites. Two sites did not consistently capture HIV diagnoses and/or diagnosis dates in structured fields. One site did not capture any HIV diagnosis dates in structured fields, and the other only captured HIV diagnoses and diagnosis dates in structured fields for patients covered by the Ryan White program. Of these two sites, one began a process of changing workflows to capture HIV diagnoses and diagnosis dates in structured fields as a result of participating in our testing efforts. Given the availability of this data element either currently or in the near term across nearly all our test sites, we do not expect this data element to substantially affect the feasibility of this measure.

3.07) Detail 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),

Attach the fee schedule here, if applicable.

Not applicable.

Use (4a.01 – 4a.10)

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

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 demonstrating performance improvement.

4a.01) Check all current uses. For each current use checked, please provide:

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

- Public Reporting
- Public Health/Disease Surveillance
- Payment Program
- Regulatory and Accreditation Programs
- Professional Certification or Recognition Program
- Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
- Quality Improvement (Internal to the specific organization)
- Not in use
- Use unknown
- Other (please specify here:)

4a.02) Check all planned uses.

- Public reporting
- Public Health/Disease Surveillance
- Payment Program
- Regulatory and Accreditation Program
- Professional Certification or Recognition Program
- Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
- Quality Improvement (internal to the specific organization)
- Measure Currently in Use
- Other (please specify here:)

4a.03) If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

This is a new eCQM that has not been used in MIPS.

4a.04) If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A 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.

HRSA plans to submit this measure for use as a clinician-level measure in the CMS MIPS program. HRSA will submit the measure to the 2023 Measures Under Consideration (MUC) list by April 30, 2023

4a.05) Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail 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.

Throughout the testing process we provided clinical practices that participated in testing with measure specifications, data dictionaries, value sets, and fact sheets to assist them in generating datasets used for testing. This measure has not yet been implemented.

4a.06) Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

Not applicable. Measure has not yet been implemented.

4a.07) Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

Not applicable. Measure has not yet been implemented.

4a.08) Summarize the feedback obtained from those being measured.

Not applicable. Measure has not yet been implemented.

4a.09) Summarize the feedback obtained from other users.

Not applicable. Measure has not yet been implemented.

4a.10) Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

Not applicable. Measure has not yet been implemented.

Usability (4b.01 - 4b.03)

4b.01) You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, 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, provide an explanation. 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.

This measure is not currently in use for performance improvement. During measure testing, clinicians at six of seven test sites agreed that the measure could be used to improve quality of care at their practices. Two of these clinicians noted that practices could use the measure to target outreach or team-based care to patients who have not been retained in care, and another noted that the measure may incentivize practices to proactively work with patients to prevent them from dropping out of care. One of the six clinicians noted that the measure may be most effective for improving quality of care when focusing on patients who are both not retained in care and not virally suppressed, and it may have more limited use when not paired with data on viral suppression. As noted above in 1a.01, retaining patients in care helps those patients achieve viral suppression, which improves patient health and reduces the risk of transmission to others. HRSA HAB is considering these measures for use in CMS's Quality Payment Programs (QPP), in particular the Merit-based Incentive Payment System (MIPS), which would provide a financial incentive for clinicians to ensure their patients are retained in care.

4b.02) Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

Not applicable. Measure has not yet been implemented.

4b.03) Explain any unexpected benefits realized from implementation of this measure.

Not applicable. Measure has not yet been implemented.

Related and Competing (5.01 - 5.06)

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01) Search and select all endorsed related measures (conceptually, either same measure focus or target population) by going to the [PQM website](#)

(Can search and select measures.)

- 3209e: HIV Medical Visit Frequency
- 3210e: HIV Viral Load Suppression
- 3211e: Prescription of HIV Antiretroviral Therapy
- 0409: HIV/AIDS: Sexually Transmitted Diseases- Screening for Chlamydia, Gonorrhea, and Syphilis (CQM only)

5.02) Search and select all endorsed competing measures (conceptually, the measures have both the same measure focus or target population) by going to the [PQM website](#).

(Can search and select measures.)

- 3209e: HIV Medical Visit Frequency

5.03) If there are related or competing measures to this measure, but they are not endorsed, please indicate the measure title and steward.

- 0410: HIV/AIDS: Sexually Transmitted Diseases - Syphilis Screening (NCQA, endorsement removed)
- 0411: HIV/AIDS: Other Infectious Diseases - Hepatitis B Screening (NCQA, endorsement removed)
- 0412: HIV/AIDS: Hepatitis B Vaccination (NCQA, endorsement removed)
- 0413: HIV/AIDS: Screening for High Risk Sexual Behaviors (NCQA, endorsement removed)
- 0414: HIV/AIDS: Other Infectious Diseases - Hepatitis C (NCQA, endorsement removed)
- 0415: HIV/AIDS: Screening for Injection Drug Use (NCQA, endorsement removed)

5.04) If this measure conceptually addresses EITHER the same measure focus OR the same target population as endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

- Yes
 No

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

This measure and 3209e are competing measures that use different approaches for measure retention in care for people with HIV. This measure defines retention based on encounters and viral load tests within a one-year period, assessing whether patients had at least two encounters or one encounter and one viral load at least 90 days apart. 3209e assesses retention over a two-year period, assessing whether patients had a medical visit at least once every six months over that period. Given that 3209e is not in use in MIPS as an eCQM, we do not expect the differences between the specifications to meaningfully affect provider data collection burden. The denominator population for this measure differs slightly from three related measures—3210e, 3211e, and 0409—with respect to the timing of the patient’s HIV diagnosis and eligible encounter, and these differences are due to the specific timing required for assessing retention in care. We limit the population to patients diagnosed no earlier than eight months into the performance period, and with an eligible encounter no later than eight months into the performance period, because our expert workgroup and the clinicians we interviewed during testing agreed that this specification allowed sufficient time for a clinician to have a follow-up encounter or viral load test 90 days after the initial encounter and before the end of the measurement period. Given that neither 3210e, 3211e, nor 0409 are currently in use in MIPS as eCQMs, we do not expect these differences to meaningfully affect data collection burden. Moreover, should CMS opt to include both this measure and 3210e in MIPS in future years, both measures would rely on encounter and HIV diagnosis dates captured in the EHR, and the differences between the measure specifications would not be expected to increase clinician burden. The measure also differs from the denominator population for 0409 with respect to the patient’s age because retention in care is a relevant process measure for all patients with HIV, regardless of age, while 0409 focuses on older patients who may be sexually active. Again, given that 0409 is not currently in use in MIPS as an eCQM, we do not expect these differences to meaningfully affect data collection burden.

5.06) Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

HIV Annual Retention in Care: This measure assesses retention in care within a one-year timeframe, which is consistent with how the HRSA HAB Ryan White HIV/AIDS Program measures retention within its annual data report. By comparison, 3209e tracks retention within a 24-month timeframe. Ryan White recipients are used to tracking retention within a one-year timeframe.

Additional (1 - 9)

1) Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

- Available in attached file
- No appendix
- Available at measure-specific web page URL identified in sp.09

2) List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

The technical expert panel (TEP) is a multi-stakeholder group with expertise in HIV clinical care, quality measurement, electronic health records (EHR), and patient and family representatives. Members of the TEP are listed in the table below. The TEP was convened three times: 1) prior to testing to provide feedback on initial measure specifications, 2) after importance and feasibility testing and the public comment period to review results and make recommendations for updates to specifications and approaches to validity and reliability testing, and 3) after validity and reliability testing to review results and evaluate the measure against NQF criteria.

The names and affiliations of the technical expertise panel members are:

Laura Bachmann, CDC

Kathleen Brady, Department of Public Health, Philadelphia

Crystal Chapman Lambert, University of Alabama

Jonathan Colasanti, Grady Hospital; Emory University

Elizabeth DiNenno, CDC

Thomas Gift, CDC

Thomas Giordano, Harris County Hospital District (Houston, TX); Baylor College of Medicine

Travis Gossey, Weill Cornell Medical College

David Harvey, National Coalition of STD Directors (NCSD)

Michael Horberg, Kaiser Permanente

Sheila Salvant Valentine, CDC

Shannon Sims, Vizient, Inc

Michelle Van Handel, CDC

Abby Viall, CDC (embedded at CMS)

Andrea Weddle, HIV Medical Association

Patient Experience Representative

3) Indicate the year the measure was first released.

N/A

4) Indicate the month and year of the most recent revision.

N/A

5) Indicate the frequency of review, or an update schedule, for this measure.

N/A

6) Indicate the next scheduled update or review of this measure.

N/A

7) Provide a copyright statement, if applicable. Otherwise, indicate “N/A”.

This measure was developed by the Health Resources and Services Administration of the U.S. Department for Health and Human Services. It is in the public domain.

Citation of HRSA as the source of the original measure is appreciated. Any modified versions may not be represented as approved, endorsed, or authorized by HRSA or HHS. 42 U.S.C. § 1320b-10. Users of modified versions should clearly explain how they deviate from HRSA’s original measure.

Limited proprietary coding is contained in the Measure specifications for user convenience. Users of proprietary code sets should obtain all necessary licenses from the owners of the code sets.

CPT(R) contained in the Measure specifications is copyright 2004-2022 American Medical Association. LOINC(R) is copyright 2004-2022 Regenstrief Institute, Inc. This material contains SNOMED Clinical Terms(R) (SNOMED CT[R]) copyright 2004-2022 International Health Terminology Standards Development Organisation. ICD-10 is copyright 2022 World Health Organization. All Rights Reserved.

Due to technical limitations, registered trademarks are indicated by (R) or [R] and unregistered trademarks are indicated by (TM) or [TM].

8) State any disclaimers, if applicable. Otherwise, indicate “N/A”.

These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.

THE MEASURES AND SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF ANY KIND.

Due to technical limitations, registered trademarks are indicated by (R) or [R] and unregistered trademarks are indicated by (TM) or [TM].

9) Provide any additional information or comments, if applicable. Otherwise, indicate “N/A”.

N/A