



## Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

### Brief Measure Information

**NQF #:** 3755e

**Corresponding Measures:**

**Measure Title:** STI Testing for People with HIV

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

**sp.02. Brief Description of Measure:** Percentage of patients 13 years of age and older with a diagnosis of HIV who had tests for syphilis, gonorrhea, and chlamydia performed within the measurement period

**1b.01. Developer Rationale:**

The rates of syphilis, gonorrhea, and chlamydia cases per 100,000 in the United States have steadily risen over the last decade and increased 11.2%, 5.9%, and 2.8%, respectively, from 2018 to 2019 (DHHS, 2021). People with HIV are at an increased risk of bacterial STIs, including chlamydia, gonorrhea, and syphilis (CDC 2004). However, early detection and treatment of bacterial STIs in people with HIV can lead to a reduction in HIV transmission (CDC 2004). Despite guidelines for at least annual screening among sexually active persons with HIV, only an estimated 55% received a syphilis test in the past year, 23% received a gonorrhea test in the past year, and 24% received a chlamydia test in the past year based on a nationally representative survey of adults with HIV receiving medical care in the United States (Flagg et al., 2015). In an analysis of people with HIV enrolled in a large integrated managed care consortium using electronic health record data, Black people with HIV were less likely to receive syphilis screening and women with HIV were less likely to receive chlamydia and gonorrhea screening (Hojilla et al., 2022). This measure will help providers focus their attention and quality improvement efforts towards testing and treating sexually transmitted infections in patients with HIV, thus reducing the complications to long-term syphilis infection and reducing STI incidence (Patel et al, 2012).

**CITATIONS:**

*Centers for Disease Control and Prevention, Health Resources and Services Administration, National Institutes of Health, HIV Medicine Association of the Infectious Diseases Society of America, & HIV Prevention in Clinical Care Working Group (2004). Recommendations for incorporating human immunodeficiency virus (HIV) prevention into the medical care of persons living with HIV. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 38(1), 104–121. <https://doi.org/10.1086/380131>*

*Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2019. (2021). Atlanta: U.S. Department of Health and Human Services. Available at: <https://www.cdc.gov/std/statistics/2019/default.htm>*

*Flagg, E. W., Weinstock, H. S., Frazier, E. L., Valverde, E. E., Heffelfinger, J. D., & Skarbinski, J. (2015). Bacterial sexually transmitted infections among HIV-infected patients in the United States: estimates from the Medical Monitoring Project. Sexually transmitted diseases, 42(4), 171–179.*

<https://doi.org/10.1097/OLQ.0000000000000260>. Erratum in: *Sex Transm Dis*. 2015 Jun;42(6):351-2. PMID: 25763669; PMCID: PMC6921480.

Hojilla, J. C., Sarovar, V., Lam, J. O., Park, I. U., Vincent, W., Hare, C. B., Silverberg, M. J., & Satre, D. D. (2022). Sexually Transmitted Infection Screening in Key Populations of Persons Living with HIV. *AIDS and Behavior*. <https://doi.org/10.1007/s10461-022-03747-w>

Patel, P., Bush, T., Mayer, K., Milam, J., Richardson, J., Hammer, J., Henry, K., Overton, T., Conley, L., Marks, G., Brooks, J. T., & SUN Study Investigators (2012). Routine brief risk-reduction counseling with biannual STD testing reduces STD incidence among HIV-infected men who have sex with men in care. *Sexually transmitted diseases*, 39(6), 470–474. <https://doi.org/10.1097/OLQ.0b013e31824b3110>

---

**sp.12. Numerator Statement:** Patients who had a test for syphilis, a test for gonorrhea, and a test for chlamydia performed at least once during the measurement period.

**sp.14. Denominator Statement:** All patients 13 years of age and older with a diagnosis of HIV before the end of the measurement period seen for an eligible encounter during 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:**

**Most Recent Endorsement Date:**

---

**IF this measure is included in a composite, NQF Composite#/title:**

**IF this measure is paired/grouped, NQF#/title:**

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

## 1. Importance to Measure and Report

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

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:**

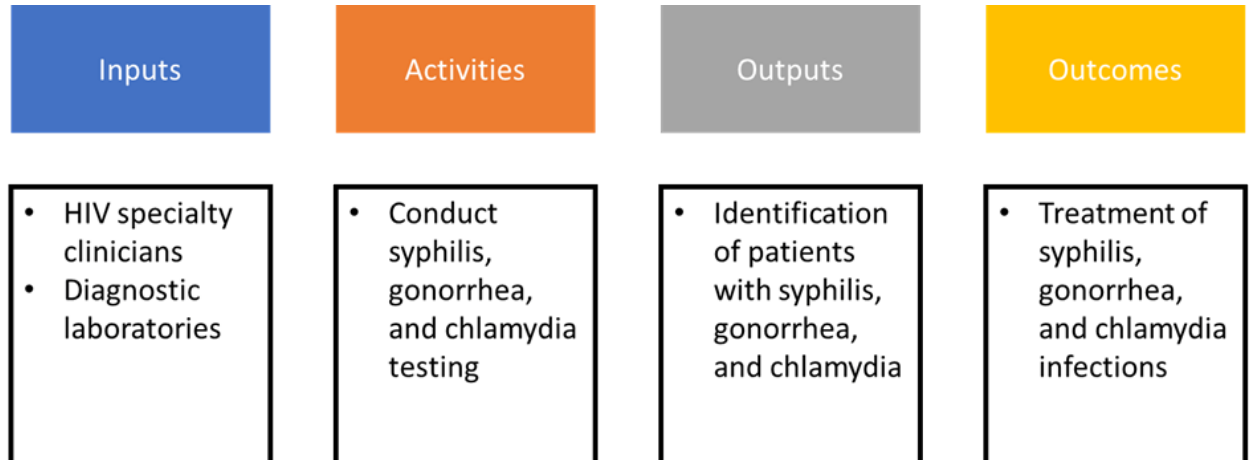
Evidence from the previous submission here.

**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.*

**[Response Begins]**

Figure 1a.01 Logic Model



Inputs to the process include HIV specialty clinicians and diagnostic laboratories. These inputs feed into the following activities: HIV specialty clinicians refer their patients to diagnostic laboratories, which conduct syphilis, gonorrhea, and chlamydia testing. These activities result in the output of identification of patients with syphilis, gonorrhea, and chlamydia (Workowski et al., 2021). This output results in the outcome of treatment of syphilis, gonorrhea, and chlamydia (Tuddenham et al., 2022).

**ALT-TEXT:**

Figure 1a.01 shows the processes of HIV primary care, including the inputs, activities, outputs, and outcome, that are related to STI testing. Inputs to the process include HIV specialty clinicians and diagnostic laboratories. These inputs feed into the following activities: HIV specialty clinicians refer their patients to diagnostic laboratories, which conduct STI tests. These activities result in the output of identification of patients with syphilis, gonorrhea, and chlamydia. This output results in the outcome of treatment of syphilis, gonorrhea, and chlamydia infections.

**CITATIONS:**

Tuddenham, S., Hamill, M. M., & Ghanem, K. G. (2022). *Diagnosis and Treatment of Sexually Transmitted Infections: A Review*. JAMA, 327(2), 161–172. <https://doi.org/10.1001/jama.2021.23487>

Workowski, K. A., Bachmann, L. H., Chan, P. A., Johnston, C. M., Muzny, C. A., Park, I., Reno, H., Zenilman, J. M., & Bolan, G. A. (2021). *Sexually Transmitted Infections Treatment Guidelines, 2021*. MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports, 70(4), 1–187. <https://doi.org/10.15585/mmwr.rr7004a1>

**[Response Ends]**

**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.*

**[Response Begins]**

Clinical Practice Guideline recommendation (with evidence review)

US Preventive Services Task Force Recommendation

**[Response Ends]**

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, add additional tables by clicking “Add” after the final question in the group.

**Evidence - Systematic Reviews Table (Repeatable)**

Group 1 - Evidence - Systematic Reviews Table

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

**[Response Begins]**

*Panel on Opportunistic Infections in Adults and Adolescents with HIV. Guidelines for the prevention and treatment of opportunistic infections in adults and adolescents with HIV: recommendations from the Centers for Disease Controls and Prevention, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. Available at: <https://clinicalinfo.hiv.gov/en/guidelines/hiv-clinical-guidelines-adult-and-adolescent-opportunistic-infections/syphilis?view=full>. Accessed November 2022. (Guideline updated December 5, 2015; reviewed July 13, 2022).*

*US Preventive Services Task Force, Mangione, C. M., Barry, M. J., Nicholson, W. K., Cabana, M., Chelmow, D., Coker, T. R., Davis, E. M., Donahue, K. E., Jaén, C. R., Kubik, M., Li, L., Ogedegbe, G., Pbert, L., Ruiz, J. M., Stevermer, J., & Wong, J. B. (2022). Screening for Syphilis Infection in Nonpregnant Adolescents and Adults: US Preventive Services Task Force Reaffirmation Recommendation Statement. JAMA, 328(12), 1243–1249. <https://doi.org/10.1001/jama.2022.15322>*

*Workowski, K. A., Bachmann, L. H., Chan, P. A., Johnston, C. M., Muzny, C. A., Park, I., Reno, H., Zenilman, J. M., & Bolan, G. A. (2021). Sexually Transmitted Infections Treatment Guidelines, 2021. MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports, 70(4), 1–187. <https://doi.org/10.15585/mmwr.rr7004a1>*

[Response Ends]

**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.**

[Response Begins]

"Routine serologic screening for syphilis is recommended at least annually for all persons with HIV infection who are sexually active, with more frequent screening (i.e., every 3–6 months) for those who have multiple or anonymous partners." (Panel on Opportunistic Infections in Adults and Adolescents with HIV, Y-3)

"Patients undergoing screening or treatment for syphilis also should be evaluated for other sexually transmitted diseases such as chlamydia and gonorrhea at anatomic sites of exposure in men and for chlamydia, gonorrhea, and trichomonas in women." (Panel on Opportunistic Infections in Adults and Adolescents with HIV, Y-3)

"The USPSTF recommends screening for syphilis in persons who are at increased risk for infection. When deciding which persons to screen for syphilis, clinicians should consider the prevalence of infection in the communities they serve, as well as other sociodemographic and behavioral factors that may be associated with increased risk of syphilis infection. For example, prevalence of syphilis is higher in men, men who have sex with men, persons with HIV infection, young adults, and persons with a history of incarceration, sex work, or military service.... Optimal screening frequency for persons who are at increased risk for syphilis infection is not well established. Men who have sex with men or persons with HIV infection may benefit from screening at least annually or more frequently (e.g., every 3 to 6 months) if they continue to be at high risk." (USPSTF 2022, pages 1244 & 1246).

"At the initial HIV care visit, providers should screen all sexually active persons for syphilis, gonorrhea, and chlamydia, and perform screening for these infections at least annually during the course of HIV care. Specific testing includes syphilis serology and [a nucleic acid amplification test] NAAT for *N. gonorrhoeae* and *C. trachomatis* at the anatomic site of exposure." (Workowski et al., 2021, page 26)

[Response Ends]

**1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.**

[Response Begins]

Panel on Opportunistic Infections in Adults and Adolescents with HIV- no grade assigned

USPSTF- Grade A

Sexually Transmitted Infections Treatment Guidelines- no grade assigned

[Response Ends]

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

[Response Begins]

**Table 1a.06-A. Rating Scheme for Recommendations from the US Preventative Task Force**

Grade	Definition
A	The USPSTF recommends the service. There is high certainty that the net benefit is substantial.
B	The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.

Grade	Definition
<b>C</b>	The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small.
<b>D</b>	The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits.
<b>I Statement</b>	The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

**ALT-TEXT:**

Table 1a.06 shows the US Preventative Service Task Force's evidence and recommendations rating scheme. The rating scheme grade and definitions are as follows: grade A, The USPSTF recommends the service. There is high certainty that the net benefit is substantial; grade B, The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial; grade C, The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small; grade D, The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits; and I statement, The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

**[Response Ends]**

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

**[Response Begins]**

Panel on Opportunistic Infections in Adults and Adolescents with HIV- no grade assigned

USPSTF- Grade A

Sexually Transmitted Infections Treatment Guidelines- no grade assigned

**[Response Ends]**

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

**[Response Begins]**

See 1a.06

**[Response Ends]**

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

**[Response Begins]**

This measure draws on a set of joint recommendations from the Centers for Disease Control and Prevention (CDC), the Health Resources and Services Administration, and the HIV Medicine Association on Guidelines for the Prevention and Treatment of Opportunistic Infections in Adults and Adolescents with HIV, as well as Sexually Transmitted Infections Treatment Guidelines from the US Department of Health and Human Services (HHS) and the CDC. The guidelines focused on adults and adolescents with HIV are based on a cohort study of MSM examining incidence of new and repeated syphilis infection (Branger et al 2009) and two additional sets of CDC guidelines (CDC 2004, CDC 2008). The STI treatment guidelines (Workowski et al., 2021) cite additional guidelines

from the Infectious Disease Society of America (Aberg et al. 2014) and a prospective, observational, multi-site cohort study (Patel et al. 2012).

**CITATIONS:**

Aberg, J. A., Gallant, J. E., Ghanem, K. G., Emmanuel, P., Zingman, B. S., Horberg, M. A., & Infectious Diseases Society of America (2014). *Primary care guidelines for the management of persons infected with HIV: 2013 update by the HIV medicine association of the Infectious Diseases Society of America. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 58(1), e1–e34. <https://doi.org/10.1093/cid/cit665>

Branger, J., van der Meer, J. T., van Ketel, R. J., Jurriaans, S., & Prins, J. M. (2009). High incidence of asymptomatic syphilis in HIV-infected MSM justifies routine screening. *Sexually transmitted diseases*, 36(2), 84–85. <https://doi.org/10.1097/OLQ.0b013e318186debb>

Centers for Disease Control and Prevention (CDC) (2008). *Recommendations for partner services programs for HIV infection, syphilis, gonorrhea, and chlamydial infection. MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports*, 57(RR-9), 1–CE4. Available at <https://pubmed.ncbi.nlm.nih.gov/18987617/>

Centers for Disease Control and Prevention, Health Resources and Services Administration, National Institutes of Health, HIV Medicine Association of the Infectious Diseases Society of America, & HIV Prevention in Clinical Care Working Group (2004). *Recommendations for incorporating human immunodeficiency virus (HIV) prevention into the medical care of persons living with HIV. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 38(1), 104–121. <https://doi.org/10.1086/380131>.

Patel, P., Bush, T., Mayer, K., Milam, J., Richardson, J., Hammer, J., Henry, K., Overton, T., Conley, L., Marks, G., Brooks, J. T., & SUN Study Investigators (2012). Routine brief risk-reduction counseling with biannual STD testing reduces STD incidence among HIV-infected men who have sex with men in care. *Sexually transmitted diseases*, 39(6), 470–474. <https://doi.org/10.1097/OLQ.0b013e31824b3110>

**[Response Ends]**

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

**[Response Begins]**

Studies, guidelines, and recommendations showed strong evidence for screening people with HIV (PWH) for syphilis and recommended co-testing for gonorrhea and chlamydia for PWH undergoing syphilis screening. The recommendations indicate that screening should occur annually for sexually active PLWH.

**[Response Ends]**

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

**[Response Begins]**

None identified.

**[Response Ends]**

**1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.**

**[Response Begins]**

N/A- the most recent guidelines are included.

**[Response Ends]**

**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.**

**[Response Begins]**

Not applicable.

**[Response Ends]**

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

**[Response Begins]**

See 1a.04 and 1a.09.

**[Response Ends]**

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

**[Response Begins]**

Used guidelines.

**[Response Ends]**

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

**[Response Begins]**

See 1a.04 and 1a.09.

**[Response Ends]**

**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.*

**[Response Begins]**

The rates of syphilis, gonorrhea, and chlamydia cases per 100,000 in the United States have steadily risen over the last decade and increased 11.2%, 5.9%, and 2.8%, respectively, from 2018 to 2019 (DHHS, 2021). People with HIV are at an increased risk of bacterial STIs, including chlamydia, gonorrhea, and syphilis (CDC 2004). However, early detection and treatment of bacterial STIs in people with HIV can lead to a reduction in HIV transmission (CDC 2004). Despite guidelines for at least annual screening among sexually active persons with HIV, only an estimated 55% received a syphilis test in the past year, 23% received a gonorrhea test in the past year, and 24% received a chlamydia test in the past year based on a nationally representative survey of adults with HIV receiving medical care in the United States (Flagg et al., 2015). In an analysis of people with HIV enrolled in a large integrated managed care consortium using electronic health record data, Black people with HIV were less to receive syphilis screening and women with HIV were less likely to receive chlamydia and gonorrhea screening (Hojilla et al., 2022). This measure will help providers focus their attention and quality improvement efforts towards testing and treating sexually transmitted infections in patients with HIV, thus reducing the complications to long-term syphilis infection and reducing STI incidence (Patel et al, 2012).

**CITATIONS:**

*Centers for Disease Control and Prevention, Health Resources and Services Administration, National Institutes of Health, HIV Medicine Association of the Infectious Diseases Society of America, & HIV Prevention in Clinical Care Working Group (2004). Recommendations for incorporating human immunodeficiency virus (HIV) prevention into*



the medical care of persons living with HIV. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 38(1), 104–121. <https://doi.org/10.1086/380131>

Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2019*. (2021). Atlanta: U.S. Department of Health and Human Services. Available at: <https://www.cdc.gov/std/statistics/2019/default.htm>

Flagg, E. W., Weinstock, H. S., Frazier, E. L., Valverde, E. E., Heffelfinger, J. D., & Skarbinski, J. (2015). Bacterial sexually transmitted infections among HIV-infected patients in the United States: estimates from the Medical Monitoring Project. *Sexually transmitted diseases*, 42(4), 171–179.

<https://doi.org/10.1097/OLQ.0000000000000260>. Erratum in: *Sex Transm Dis*. 2015 Jun;42(6):351-2. PMID: 25763669; PMCID: PMC6921480.

Hojilla, J. C., Sarovar, V., Lam, J. O., Park, I. U., Vincent, W., Hare, C. B., Silverberg, M. J., & Satre, D. D. (2022). Sexually Transmitted Infection Screening in Key Populations of Persons Living with HIV. *AIDS and Behavior*. <https://doi.org/10.1007/s10461-022-03747-w>

Patel, P., Bush, T., Mayer, K., Milam, J., Richardson, J., Hammer, J., Henry, K., Overton, T., Conley, L., Marks, G., Brooks, J. T., & SUN Study Investigators (2012). Routine brief risk-reduction counseling with biannual STD testing reduces STD incidence among HIV-infected men who have sex with men in care. *Sexually transmitted diseases*, 39(6), 470–474. <https://doi.org/10.1097/OLQ.0b013e31824b3110>

### [Response Ends]

#### 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.

### [Response Begins]

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 123 clinicians and 2,990 patients participating in the Ryan White HIV/AIDS Program, of which 37 clinicians had at least 11 patients—the minimum sample size requirement outlined in the CMS cell suppression policy. The data for these 37 clinicians included 2,891 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 denominator-eligible patients (Table 1b.02).

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

	Clinicians	Patients	Mean	SD	Min	p10	p20	p30	p40	Median	p60	p70	p80	p90	Max	IQR
All Clinicians	123	2,990	55.9	36.7	0.0	0.0	10.3	40.1	50.0	60.0	72.2	80.0	100.0	100.0	100.0	76.9
Clinicians with 11+ denominator-eligible patients	37	2,891	54.5	24.2	10.9	14.6	32.1	44.4	52.6	60.3	65.3	72.2	78.6	80.1	95.1	54.5

Notes: SD=Standard deviation, Min=minimum, Max=maximum, p=percentile, IQR=interquartile range.

**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.

[Response Ends]

**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.**

[Response Begins]

Not applicable.

[Response Ends]

**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.*

[Response Begins]

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 STI testing 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 STI Testing across clinicians with ≥11 patients eligible for the denominator (n = 37)**

Patient Group	Clinicians	Patients	Mean	Std dev	Min	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	Max	IQ R
AGE<50	37	1,456	61.3	25.1	17.6	23.5	33.6	45.9	56.2	66.7	74.3	80.4	85.4	87.2	100.0	45.6
AGE≥50	37	1,435	45.5	25.4	3.6	6.7	22.1	33.3	40.0	42.9	53.0	63.3	70.5	76.0	93.8	38.9
MSM	34	1,179	60.1	30.2	0.0	16.5	26.0	44.8	58.7	65.2	75.0	81.5	87.8	95.4	100.0	50.5
IDU	30	193	41.6	35.4	0.0	0.0	0.0	11.0	28.0	36.7	52.9	62.0	72.1	100.0	100.0	59.3

Patient Group	Clinicians	Patients	Mean	Standard deviation	Min	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	Max	IQR
Other transmission	34	919	45.6	26.2	5.4	10.3	21.1	32.6	38.0	42.2	49.8	62.6	66.7	74.3	100.0	40.0
Black	37	1,270	55.7	24.1	14.3	22.1	27.3	42.3	51.8	60.0	64.7	71.1	74.7	83.3	100.0	31.7
White	37	1,496	53.7	29.3	0.0	11.2	21.7	36.7	50.0	58.8	65.9	71.4	76.0	91.2	100.0	41.1
Not Hispanic or Latino	37	2,443	53.3	24.3	11.1	14.3	31.6	40.0	50.6	56.7	63.8	71.6	75.7	81.3	92.3	34.5
Hispanic or Latino	27	440	57.0	33.2	0.0	10.7	15.6	41.0	51.2	66.7	75.0	82.0	83.6	98.1	100.0	63.1

Notes: Results are for clinicians with  $\geq 11$  patients eligible for the denominator.

SD=standard deviation, p=percentile, 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 under fifty years, patients with non-IDU or MSM transmission methods, White patients and Hispanic or Latino patients.

Rates of STI testing were higher among those under age 50 (mean=61.3%) as compared to those age 50 and older (mean=45.5%). The differences in the measure score by patients' age were statistically significant ( $p=0.01$ ), which reflects higher rates of testing among younger patients (Berry et al 2015). STI testing rates were higher among MSM (mean=60.1%) than other HIV transmission groups (41.6% for IDU, 45.6% for other transmission) [ $p$ -value=0.06], as expected based on the literature (Berry et al 2015), although the differences did not reach statistical significance at the  $p \leq 0.05$  level. Rates were similar across race ( $p$ -value=0.63) and ethnicity ( $p$ -value=0.74).

Across all clinicians, we observed more variability in the clinician-level measure rates for the white patients (mean=53.7, SD=29.3) patients, and the IDU patients (mean=41.6, SD=35.4). Larger variation in the measure rates for the IDU patients can be attributed to the relatively small sample for that category, whereas variability in the measure rates for the white patients can point at the potential differences in care within this group of patients or differences in the distribution of other characteristics that might influence STI screening, e.g. age or perceived risk.

#### CITATIONS:

Berry, S. A., Ghanem, K. G., Mathews, W. C., Korthuis, P. T., Yehia, B. R., Agwu, A. L., Lehmann, C. U., Moore, R. D., Allen, S. L., Gebo, K. A., & HIV Research Network (2015). Brief Report: Gonorrhea and Chlamydia Testing Increasing but Still Lagging in HIV Clinics in the United States. *Journal of acquired immune deficiency syndromes (1999)*, 70(3), 275–279. <https://doi.org/10.1097/QAI.0000000000000711>

[Response Ends]

**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.**

**[Response Begins]**

[Not applicable.](#)

**[Response Ends]**

## 2. Scientific Acceptability of Measure Properties

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

---

### sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see [What Good Looks Like](#)).

#### [Response Begins]

STI Testing for People with HIV

#### [Response Ends]

### 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).

#### [Response Begins]

Percentage of patients 13 years of age and older with a diagnosis of HIV who had tests for syphilis, gonorrhea, and chlamydia performed within the measurement period

#### [Response Ends]

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

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- Surgery: General

#### [Response Begins]

Infectious Diseases (ID)

Infectious Diseases (ID): HIV/AIDS

Infectious Diseases (ID): Sexually Transmitted

#### [Response Ends]

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

#### [Response Begins]

Screening

#### [Response Ends]

### 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.

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Populations at Risk: Populations at Risk*

**[Response Begins]**

Adults (Age >= 18)

Children (Age < 18)

**[Response Ends]**

**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.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Clinician: Clinician*
- *Population: Population*

**[Response Begins]**

Clinician: Individual

**[Response Ends]**

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

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

**[Response Begins]**

Inpatient/Hospital

Outpatient Services

**[Response Ends]**

**sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.**

*Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".*

**[Response Begins]**

[None available.](#)

**[Response Ends]**

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

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

**[Response Begins]**

HQMF specifications are attached.

**[Response Ends]**

Attachment: 3755e\_CMS1188-v0-0-009-QDM-5-6.zip

**sp.11. Attach the simulated testing attachment.**

*All eQMs 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.*

**[Response Begins]**

Testing is attached

**[Response Ends]**

Attachment: 3755e\_CMS1188v0.xlsx

**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](#). Provide descriptors for any codes. Use one file with multiple worksheets, if needed.*

**[Response Begins]**

Available in attached Excel or csv file

**[Response Ends]**

Attachment: 3755e\_CMS1188+HIV+STI+Test+(2022-10-31+16-13-37).xlsx

**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.*

**[Response Begins]**

Patients who had a test for syphilis, a test for gonorrhea, and a test for chlamydia performed at least once during the measurement period.

**[Response Ends]**

**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.*

**[Response Begins]**

Qualifying syphilis, gonorrhea, and chlamydia 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.

**[Response Ends]**

**sp.15. State the denominator.**

*Brief, narrative description of the target population being measured.*

**[Response Begins]**

All patients 13 years of age and older with a diagnosis of HIV before the end of the measurement period seen for an eligible encounter during the measurement period.

**[Response Ends]**

**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.*

**[Response Begins]**

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 eligible encounter can occur at any point in the measurement period because patients should receive screening at least annually, including at their first visit with a provider.

**[Response Ends]**

**sp.17. Describe the denominator exclusions.**

*Brief narrative description of exclusions from the target population.*

**[Response Begins]**

Not applicable.

**[Response Ends]**

**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.*



[Response Begins]

Not applicable.

[Response Ends]

**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.*

[Response Begins]

Not applicable.

[Response Ends]

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

[Response Begins]

No

[Response Ends]

**sp.21. Select the risk adjustment type.**

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

[Response Begins]

No risk adjustment or risk stratification

[Response Ends]

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

*Attachment: If available, please provide a sample report.*

[Response Begins]

Rate/proportion

[Response Ends]

**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*

[Response Begins]

Better quality = Higher score

[Response Ends]

**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.*

**[Response Begins]**

**Denominator**

1. Identify patients with an eligible encounter during the measurement period
2. Retain all patients diagnosed with HIV during the measurement period or any time prior.
3. Retain all patients 13 years or older.
4. Patients meeting these criteria are in the denominator, and those that do not meet these criteria are not in the denominator.

**Numerator**

1. Identify denominator-eligible patients with a syphilis test, a gonorrhea test, and a chlamydia test during the measurement period.
2. If the patient has all three tests during the measurement period, the patient is included in the numerator. If the patient does not have all three tests during the measurement period, the patient is not included in the numerator.

**[Response Ends]**

**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 [2010 Measure Testing Task Force](#) recognized that 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.

**[Response Begins]**

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 37 clinicians and 2,891 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 37 clinicians. As a part of reliability testing 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)

**[Response Ends]**

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

**[Response Begins]**

Electronic Health Records

**[Response Ends]**

**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.*

**[Response Begins]**

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

**[Response Ends]**

**sp.32. Provide the data collection instrument.**

**[Response Begins]**

No data collection instrument provided

**[Response Ends]**

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 NQF staff 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 NQF staff with any questions. Check for resources at the [Submitting Standards webpage](#).
- 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 NQF's 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
- 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 measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. 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.

#### [Response Begins]

Electronic Health Records

#### [Response Ends]

### 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).*

**[Response Begins]**

This submission uses the patient-level data from seven participating sites. Please refer to section sp.27 for more details about the sample.

**[Response Ends]**

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

*Use the following format: “MM-DD-YYYY - MM-DD-YYYY”*

**[Response Begins]**

01-01-2021 – 12-31-2021

**[Response Ends]**

**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.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- Clinician: Clinician
- Population: Population

**[Response Begins]**

Clinician: Individual

**[Response Ends]**

**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.*

**[Response Begins]**

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 37 clinicians were included in testing. These 37 clinicians had a total of 2,891 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	Clinicians with 11+ patients	Patients
Site 1	Publicly funded community health center	NE	eClinical Works	2	97
Site 2	Publicly funded community health center	NE	eClinical Works	4	162
Site 3	Hospital or university-based clinic	NE	EPIC	16	394
Site 4	Publicly funded community health center	SO	OCHIN EPIC	3	574
Site 5	Other community-based service organization	NE	NextGen	3	51
Site 6	Publicly funded community health center	MW	Athena Health	3	560
Site 7	Hospital or university-based clinic	MW	eClinical Works	6	1,053

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.

#### **[Response Ends]**

**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.*

#### **[Response Begins]**

Data included 2,990 patients attributed to 123 clinicians within the measurement period from 7 different 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, in order 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. The restricted sample includes 37 clinicians (30.1% of the initial number of clinicians) and 2,891 patients (96.7% of the initial number of patients). When limited to clinicians with 11 or more patients eligible for the denominator during the measurement period, the average (mean) clinician has an STI Testing measure rate of 54.5%.

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

- 3% (98) of patients were under the age of 25 and 97% (2,793) of patients were over the age of 25
- 50% (1,456) of patients were under the age of 50, and 50% (1,435) of patients were over the age of 50.
- Broken out by HIV transmission group, 41% (1,179) of patients' transmission group was men who have sex with men (MSM), 7% (193) 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 32% (919) of patients' transmission group was other, while 21% (600) of patients were missing information on HIV transmission group.

- 25% (711) patients were cisgender women, 75% (2,165) patients were cisgender men, and 1% (15) patients were transgender women.

**[Response Ends]**

**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.**

**[Response Begins]**

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 (MSM vs. non-MSM). 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 an email poll of the Technical Expert Panel (TEP) that was convened during the development of the STI Testing for People with HIV 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 there were 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.

**[Response Ends]**

**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.*

**[Response Begins]**

Not applicable. This is a process measure.

**[Response Ends]**



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.

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

*Choose one or both levels.*

**[Response Begins]**

Accountable Entity Level (e.g., signal-to-noise analysis)

**[Response Ends]**

**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.*

**[Response Begins]**

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 STI Testing for People with HIV 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](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](https://www.qualityforum.org/Projects/n-r/Patient-Reported_Outcomes/Commissioned_Paper_2.aspx)

**[Response Ends]**

#### **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, [NQF Measure Evaluation Criteria](#)).*

**[Response Begins]**

Table 2a.11-A summarizes the mean and range of the signal-to-noise reliability statistics for the STI Testing measure, which was calculated separately for each clinician. The mean signal-to-noise reliability across all 37 clinicians with at least 11 patients eligible for the denominator exceeds the 0.70 threshold for acceptable reliability. The 25th percentile for the measure reliability was 0.774, and the 75th percentile was 0.962.

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

Sample	Clinicians	Min	Mean	SD	p5	p10	p25	p50	p75	p90	p95	Max
Clinicians >= 11 patients	37	0.612	0.864	0.107	0.710	0.731	0.774	0.891	0.962	0.987	0.988	0.989

Sample	Clinicians	Min	Mean	SD	p5	p10	p25	p50	p75	p90	p95	Max
Clinicians >= 11 patients	37	0.612	0.864	0.107	0.710	0.731	0.774	0.891	0.962	0.987	0.988	0.989

Notes: Results are provided for clinicians with  $\geq 11$  patients eligible for the denominator (n = 37). Min=minimum, SD=standard deviation, Max=maximum

#### 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 STI Testing measure. All split-half reliability correlations exceed 0.7 with a range from 0.844 to 0.966.

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

Sample	Spearman rank-order correlation	Spearman-Brown correlation	intra-class correlation
Clinicians >= 11 patients	0.934	0.966	0.844

Notes: Results are provided for clinicians with  $\geq 11$  patients eligible for the denominator (n = 37)

#### 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 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 STI Testing measure, which was calculated separately by each sample. The mean test-retest reliability ranges from 0.911 to 0.978 with the 25th percentile ranging from 0.890 to 0.974.

**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	0.958	0.829	0.933	0.939	0.950	0.960	0.968	0.975	0.978	0.989
Clinicians >= 11 patients	Spearman-Brown correction	0.978	0.906	0.965	0.968	0.974	0.979	0.984	0.987	0.989	0.995
Clinicians >= 11 patients	intra-class correlation	0.911	0.652	0.826	0.851	0.890	0.920	0.944	0.957	0.965	0.987

Notes: Results are provided for clinicians with  $\geq 11$  patients eligible for the denominator (n = 37)

#### 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.

**[Response Ends]**

**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?)*

**[Response Begins]**

The STI Testing for People with HIV 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 that 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 with at least 11 patients in the denominator 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 Portney 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 0.90 indicate good reliability, and values greater than 0.90 indicate excellent reliability. The ICCs exceeding 0.75 and 0.90 obtained from the split-half and test-retest testing methods, respectively, indicate good to 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 STI Testing 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](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](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>.

[Response Ends]

**2b.01. Select the level of validity testing that was conducted.**

[Response Begins]

Patient or Encounter-Level (data element validity must address ALL critical data elements)

Empirical validity testing

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

[Response Ends]

**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.*

[Response Begins]

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 eQCM.

**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. We tested data element validity by examining the relationship between the data elements and the measure components from two sources for a randomly selected set of patients:

1. EHR data extracted from the structured fields used by the eQCMs
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 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 for each key data element. 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 eQCMs. 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 an email poll of the Technical Expert Panel (TEP) that was convened during the development of the STI Testing for

People with HIV 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 STI testing among people with HIV: transmission category (men who have sex with men [MSM] versus non-MSM) 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, calculated the measure rates, and computed 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>

**[Response Ends]**

#### **2b.03. Provide the statistical results from validity testing.**

*Examples may include correlations or t-test results.*

**[Response Begins]**

**Data element validity.** We found that the percent agreement between the EHR data and manual abstraction ranged across the measure's data elements 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 (140 records collected from 7 sites)**

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
Syphilis test confirmation	99.3	0.99
Gonorrhea test confirmation	100.0	1.00
Chlamydia test confirmation	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 6/7 (86%) of clinicians supported the measure's numerator and denominator specifications, and 3/7 (43%) of clinicians agreed that the measure can distinguish quality of care. We also found that 100% of the Technical Expert Panel (12 out of 12) agreed that the STI Testing for People with HIV 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 STI testing was 61.3% for those under 50 years as compared to 45.5% for those 50 years or older, with an effect size of 0.629 using Cohen's D. The mean rate of STI testing was 60.1 % for those whose HIV transmission group was MSM as compared to 45.7% for those whose HIV transmission group was not MSM, with an effect size of -0.525 using Cohen's D, indicating a moderate effect.

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

Patient sub-group	Mean STI Testing rates	S.D.	Cohen's D	t-test p-value
Age < 50 years	61.3	25.1	0.629	0.008
Age >= 50 years	45.5	25.4		
Non-MSM HIV transmission category	45.7	24.3	0.525	0.034
MSM HIV transmission category	60.1	30.2		

Note: 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 STI testing 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.

**[Response Ends]**

**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?)**

**[Response Begins]**

We tested validity of the measure using both qualitative (face validity) and quantitative methods (data element level: data element validity; measure score level: concurrent and known-group validity).

**Face validity.** The subject matter experts at the test sites were mixed as to whether the measure could be used to distinguish good from poor quality of care because the measure denominator is not targeted to patients who were sexually active. Their concern is that patients who are not sexually active would opt out of screening; however, sexual activity and sexual history are not well-documented in EHR in structured fields, which precludes the inclusion of them in the measure specifications. Despite this limitation, the majority supported the measure's numerator and denominator definition, indicating that despite the limitations with sexual activity structured fields, the current measure specifications are acceptable. Further, the Technical Expert Panel unanimously agreed that the measure was important and related to quality of care despite the limitations with the availability of sexual history in a structured field.

**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 95.3 percent. These statistics indicate very

high data element validity. Observed agreement was very high (above 99 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 the patient would have been included in the denominator because all the dates occurred prior to the measurement period

**Construct validity.** Our results indicated that MSM had rates of STI screening that were over 14 percentage points higher than non-MSM (0.601 vs. 0.457) and that those under age 50 had rates of STI screening that were almost 16 percentage points higher than those over 50 (0.613 vs. 0.455). The differences in these rates are similar to the hypothesized differences based on the literature, supporting the validity for this measure (Berry et al 2015). Based on a multisite HIV clinical cohort, overall testing rates were 77% for syphilis and 39% for chlamydia and gonorrhea as of 2010 with gaps in testing for non-MSM and older age groups, similar to what was observed in these results (Berry et al 2015). It should be noted that the hypothesized difference based on the literature represents a gap in care that this measure targets, namely increasing rates of STI screening among non-MSM with HIV and older age groups.

**CITATIONS:**

Berry, S. A., Ghanem, K. G., Mathews, W. C., Korthuis, P. T., Yehia, B. R., Agwu, A. L., Lehmann, C. U., Moore, R. D., Allen, S. L., Gebo, K. A., & HIV Research Network (2015). Brief Report: Gonorrhea and Chlamydia Testing Increasing but Still Lagging in HIV Clinics in the United States. *Journal of acquired immune deficiency syndromes (1999)*, 70(3), 275–279. <https://doi.org/10.1097/QAI.0000000000000711>

**[Response Ends]**

**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.*

**[Response Begins]**

To examine differences in performance, we calculated measure rates across 37 clinicians with at least 11 patients eligible for the denominator in the performance period. We excluded clinicians with less than 11 patients eligible for the denominator to comply with the CMS minimum cell size policy. We computed a confidence interval for each clinician's rate, and if it did not contain the mean rate across all clinicians, the clinician was identified as better or worse than average.

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.

**[Response Ends]**



**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.*

**[Response Begins]**

Based on 37 clinicians with at least 11 patients eligible for denominator, the STI Testing measure rates in our sample ranged from 10.9% to 95.1% (with a median of 60.3% and a mean of 54.5%). Thus, there is substantial variation in measure scores across facilities.

**Table 2b.06-A. Performance Distribution of the STI Testing measure rates**

Measure	Clinicians	Mean	Std dev	Min	10th Pctl	Lower Quartile	Median	Upper Quartile	90th Pctl	Max
clinicians 11+ patients	37	54.5%	24.2%	10.9%	14.6%	40.2%	60.3%	76.5%	80.1%	95.1%

Notes: Pctl= Percentile, Std dev= standard deviation

**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 37 clinicians, 24.3% (N=9) were statistically significantly worse and 32.4% (N = 12) were better 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 about two-thirds of clinicians whose performance scores were either no different from the sample average (42.2%) or worse than the sample average (24.3%).

**Table 2b.06-B. Performance Distribution of the STI Testing 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	12 (32.4%)	79.1%
No different than the national rate	16 (42.2%)	55.8%
Worse than the national rate	9 (24.3%)	19.5%
All clinicians	37 (100.0%)	54.5%

**ALT-TEXT:**

Table 2b.06-B shows performance distribution of the STI Testing for People with HIV 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 within two standard deviations of the 90th percentile, which meets the first criterion for being topped out; however, the measure has a truncated coefficient of variation (TCV) equal to 0.39, which does not meet the second criterion for being topped out. Thus the measure is not considered topped-out.

**Table 2b.06-C. Topped out analysis of the STI Testing for People with HIV measure rates.**

Measure	75 <sup>th</sup> pctl.	90 <sup>th</sup> pctl.	90 <sup>th</sup> – 75 <sup>th</sup> pctl.	2x S.D. of 90 <sup>th</sup> pctl.	Criterion 1 met?	TCV	Criterion 2 met?
STI Testing for People with HIV	0.76	0.80	0.04	0.02	Yes	0.39	No

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

#### **ALT-TEXT:**

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.

#### **[Response Ends]**

**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?*

#### **[Response Begins]**

There was substantial variability in the measure rates across clinicians in our sample, and the measure was able to distinguish between clinicians with better and worse than average performance scores. As only about 34 percent of clinicians had significantly better measure scores than the sample average, this indicates potential for performance improvement for nearly three quarters of clinicians in our sample whose measure scores were either worse than or not significantly different from the sample average. The measure is also not considered topped out based on testing criteria.

#### **[Response Ends]**

**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.*

#### **[Response Begins]**

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. However, the measure logic does not allow for the missing values for the encounter type and dates, as well as the patients' age, as these elements are required for the measure calculation. Therefore, we assessed the frequency of missing data elements in these fields.

#### **[Response Ends]**

**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).*

**[Response Begins]**

0% (0) patients were missing age.

0.1% (13) of the encounters had a performance date but were missing the encounter code to indicate type of encounter. This represents a very small number of the total encounters.

**[Response Ends]**

**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.*

**[Response Begins]**

Our analysis indicates that missing data are not a threat to validity for the measure due to the extremely low prevalence of missing data elements in the fields which are required for the measure calculations.

**[Response Ends]**

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.

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

**[Response Begins]**

No, there is only one set of specifications for this measure

**[Response Ends]**

**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.*

**[Response Begins]**

**[Response Ends]**

**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.*

[Response Begins]

[Response Ends]

**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.*

[Response Begins]

[Response Ends]

**2b.15. Indicate whether the measure uses exclusions.**

[Response Begins]

N/A or no exclusions

[Response Ends]

**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?*

[Response Begins]

Not applicable.

[Response Ends]

**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.*

[Response Begins]

Not applicable.

[Response Ends]

**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.*

[Response Begins]

Not applicable.

[Response Ends]

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

[Response Begins]

No risk adjustment or risk stratification

[Response Ends]

**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.**

[Response Begins]

[Response Ends]

**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.**

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

**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).*

[Response Begins]

[Response Ends]

**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.**

[Response Begins]

[Response Ends]

**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.*

[Response Begins]

[Response Ends]

**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.*

[Response Begins]

[Response Ends]

**2b.27. Provide risk model discrimination statistics.**

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

[Response Begins]

[Response Ends]

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

[Response Begins]

Not applicable.

[Response Ends]

**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.*

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

**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?*

**[Response Begins]**

**[Response Ends]**

**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.*

**[Response Begins]**

**[Response Ends]**

### 3. Feasibility

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

**3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.**

**[Response Begins]**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)

**[Response Ends]**

**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.*

**[Response Begins]**

ALL data elements are in defined fields in electronic health records (EHRs)

**[Response Ends]**

**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.**

**[Response Begins]**

All data elements are in defined fields in EHR.

**[Response Ends]**

**3.05. Complete and attach the [NQF Feasibility Score Card](#).**

**[Response Begins]**

**[Response Ends]**

Attachment: 3755e\_11082022\_STIFeasibilityScorecard.xlsx

**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.**

**[Response Begins]**

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.



**[Response Ends]**

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

**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.**

**[Response Begins]**

[Not applicable.](#)

**[Response Ends]**

## 4. Usability and Use

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

---

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.

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to 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

[Response Begins]

Not in use

[Not in use Please Explain]

Newly developed eCQM.

[Response Ends]

### 4a.02. Check all planned uses.

[Response Begins]

Payment Program

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

[Response Ends]

### 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?*

[Response Begins]

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

[Response Ends]

**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.*

**[Response Begins]**

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.

**[Response Ends]**

**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.*

**[Response Begins]**

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.

**[Response Ends]**

**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.**

**[Response Begins]**

Not applicable. Measure has not yet been implemented.

**[Response Ends]**

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

**[Response Begins]**

Not applicable. Measure has not yet been implemented.

**[Response Ends]**

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

**[Response Begins]**

Not applicable. Measure has not yet been implemented.

**[Response Ends]**

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

**[Response Begins]**

Not applicable. Measure has not yet been implemented.

[Response Ends]

**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.**

[Response Begins]

Not applicable. Measure has not yet been implemented.

[Response Ends]

**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.**

[Response Begins]

This measure is not currently in use for performance improvement. However, this measure could be used to improve quality of care by incentivizing practices to prioritize testing their patients with HIV for syphilis, gonorrhea, and chlamydia, resulting in earlier detection and treatment, particularly for asymptomatic cases. 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 receive STI testing annually.

[Response Ends]

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

[Response Begins]

Not applicable. Measure has not yet been implemented.

[Response Ends]

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

[Response Begins]

Not applicable. Measure has not yet been implemented.

[Response Ends]

## 5. Comparison to Related or Competing Measures

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

---

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 NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

### 5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

**NOTE: If there are no related measures, please select N/A.**

*(Can search and select measures.)*

#### [Response Begins]

3209e: HIV medical visit frequency

3210e: HIV viral suppression

3211e: Prescription of HIV Antiretroviral Therapy

0409: HIV/AIDS: Sexually Transmitted Diseases – Screening for Chlamydia, Gonorrhea, and Syphilis

#### [Response Ends]

### 5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus and target population).

**NOTE: If there are no competing measures, please select N/A.**

*(Can search and select measures.)*

#### [Response Begins]

0409: HIV/AIDS: Sexually Transmitted Diseases – Screening for Chlamydia, Gonorrhea, and Syphilis

#### [Response Ends]

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

#### [Response Begins]

- 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)

#### [Response Ends]

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

**[Response Begins]**

Yes

**[Response Ends]**

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

**[Response Begins]**

This measure and 0409 are competing measures, but this measure is an eCQM, while 0409 is a CQM. HRSA HAB stewards both measures (we obtained stewardship of 0409 from NCQA), and we intend to let endorsement of 0409 lapse as we seek endorsement of this measure. Consequently, we do not expect the differences between these measures to meaningfully affect provider reporting burden. The denominator population for this measure differs slightly from three related measures—3209e, 3210e, and 3211e—with respect to the timing of the patient's HIV diagnosis and eligible encounter and the patient's age, and these differences are due to the specific timing required for assessing appropriate provision of STI testing. We include patients diagnosed with HIV at any time during or prior to the measurement year and with an eligible encounter at any point during the measurement year because these parameters are consistent with the recommendations of our technical expert panel and clinicians interviewed during testing. Moreover, we limit the measure population to patients 13 years of age and older as a rough proxy for patients who may be sexually active. Given that neither 3209e, 3210e, or 3211e are currently in use in MIPS as eCQMs, we do not expect these differences to meaningfully affect data collection burden.

**[Response Ends]**

**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.*

**[Response Begins]**

STI Testing for Patients with HIV: HRSA HAB plans to let endorsement lapse for 0409, and to use the present measure as the primary vehicle for measuring appropriate provision of STI testing among patients with HIV.

**[Response Ends]**

## Appendix

**Supplemental materials may be provided in an appendix.:**

No appendix

## Contact Information

**Measure Steward (Intellectual Property Owner):** Health Resources and Services Administration - HIV/AIDS Bureau

**Measure Steward Point of Contact:** Matthews, Tracy, tmatthews@hrsa.gov

**Measure Developer if different from Measure Steward:** Health Resources and Services Administration - HIV/AIDS Bureau

**Measure Developer Point(s) of Contact:** Matthews, Tracy, tmatthews@hrsa.gov

## Additional Information

**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.**

**[Response Begins]**

No appendix

**[Response Ends]**

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

*Describe the members' role in measure development.*

**[Response Begins]**

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

**[Response Ends]**

**3. Indicate the year the measure was first released.**

**[Response Begins]**

Not applicable.



**[Response Ends]**

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

**[Response Begins]**

Not applicable.

**[Response Ends]**

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

**[Response Begins]**

Not applicable.

**[Response Ends]**

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

**[Response Begins]**

Not applicable.

**[Response Ends]**

**7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".**

**[Response Begins]**

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].

**[Response Ends]**

**8. State any disclaimers, if applicable. Otherwise, indicate "N/A".**

**[Response Begins]**

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].

**[Response Ends]**

**9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".**

**[Response Begins]**

Not applicable.

**[Response Ends]**