



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 #: 3210e

Corresponding Measures:

Measure Title: HIV viral suppression

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

sp.02. Brief Description of Measure: Percentage of patients, regardless of age, diagnosed with HIV prior to or during the first three months of the measurement period, with an eligible encounter in the first eight months of the measurement period, who have a last HIV viral load test has result of less than 200 copies/mL during the measurement period.

1b.01. Developer Rationale:

Current Submission:

HIV is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 40,000 persons in the United States are newly infected with HIV each year (Centers for Disease Control and Prevention, 2021, p. 51). Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection.

HIV viral suppression is a long-standing priority outcome among the HIV community in the United States and around the world. The National HIV/AIDS Strategy for the United States from 2022-2025, developed by the White House Office of National AIDS Policy with input from the HIV community across the United States, prioritizes increasing HIV viral suppression rates to 95% (The White House 2020). The DHHS Panel on Antiretroviral Guidelines for Adults and Adolescents defines viral suppression as a viral load below the lower limits of detection in its guidelines on virologic failure, and it defines viral suppression as a viral load of less than 200 copies/mL as part of its guidelines for the use of antiretroviral therapy to prevent HIV transmission (Panel on Antiretroviral Guidelines for Adults and Adolescents 2022).

Antiretroviral therapy (ART) delays the progression to AIDS and increases the length of survival. ART reduces HIV-associated morbidity and mortality by maximally inhibiting HIV replication to achieve viral suppression (Hogg et al 2001; Lundgren et al., 2015). ART has also been shown to reduce transmission of HIV (Rodger et al 2019). Studies show disparities in rates of viral suppression by race and ethnicity among MSM and among women, with Black and Hispanic or Latino/a study participants having lower rates of viral suppression than White participants (Buchacz et al. 2020; Buchacz et al. 2018; Geter et al. 2018). This measure will help providers direct their attention and quality improvement efforts towards improving HIV viral suppression rates.

CITATIONS:

Buchacz, K., Armon, C., Palella, F. J., Novak, R. M., Fuhrer, J., Tedaldi, E., . . . Investigators, H. O. S. H. (2020). *The HIV Outpatient Study-25 Years of HIV Patient Care and Epidemiologic Research*. *Open Forum Infect Dis*, 7(5), ofaa123. <https://doi.org/10.1093/ofid/ofaa123>

Buchacz, K., Armon, C., Tedaldi, E., Palella, F. J., Novak, R. M., Ward, D., . . . Investigators, H. O. S. (2018). *Disparities in HIV Viral Load Suppression by Race/Ethnicity Among Men Who Have Sex with Men in the HIV Outpatient Study*. *AIDS Res Hum Retroviruses*, 34(4), 357-364. <https://doi.org/10.1089/AID.2017.0162>

Centers for Disease Control and Prevention (2021). *HIV Surveillance Report, 2019*. <http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html>

Geter, A., Sutton, M. Y., Armon, C., Durham, M. D., Palella, F. J., Tedaldi, E., . . . Investigators, H. O. S. (2018). *Trends of racial and ethnic disparities in virologic suppression among women in the HIV Outpatient Study, USA, 2010-2015*. *PLoS One*, 13(1), e0189973. <https://doi.org/10.1371/journal.pone.0189973>

Hogg, R. S., Yip, B., Chan, K. J., Wood, E., Craib, K. J., O'Shaughnessy, M. V., & Montaner, J. S. (2001). *Rates of disease progression by baseline CD4 cell count and viral load after initiating triple-drug therapy*. *JAMA*, 286(20), 2568-2577. <https://doi.org/10.1001/jama.286.20.2568>

Lundgren, J. D., Babiker, A. G., Gordin, F., Emery, S., Grund, B., Sharma, S., . . . Group, I. S. S. (2015a). *Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection*. *N Engl J Med*, 373(9), 795-807. <https://doi.org/10.1056/NEJMoa1506816>

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

Rodger, A. J., Cambiano, V., Bruun, T., Vernazza, P., Collins, S., Degen, O., . . . Group, P. S. (2019). *Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study*. *Lancet*, 393(10189), 2428-2438. [https://doi.org/10.1016/S0140-6736\(19\)30418-0](https://doi.org/10.1016/S0140-6736(19)30418-0)

The White House. (2021). *National HIV/AIDS Strategy for the United States 2022–2025*. <https://files.hiv.gov/s3fs-public/NHAS-2022-2025.pdf>

Previous Submission:

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

Antiretroviral therapy reduces HIV-associated morbidity and mortality by maximally inhibiting HIV replication (as defined by achieving and maintaining plasma HIV RNA (viral load) below levels detectable by commercially available assays). Viral suppression is a main goal of HIV treatment and an indicator of treatment success and reduction of potential HIV transmission. It is directly related to:

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

sp.12. Numerator Statement: Patients with a last HIV viral load test result of less than 200 copies/ml during the measurement period

sp.14. Denominator Statement: All patients, regardless of age, diagnosed with HIV prior to or during the first three months of the measurement period with at least one eligible encounter in the first eight months of the measurement period.

sp.16. Denominator Exclusions: Not applicable.

Measure Type: Outcome

sp.28. Data Source:

Other

Electronic Health Records

sp.07. Level of Analysis:

Clinician: Individual

IF Endorsement Maintenance – Original Endorsement Date: 2017-07-13 05:46 PM

Most Recent Endorsement Date: 7/13/2017 5:46:52 PM

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

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

[Response Begins]

Yes

[Yes Please Explain]

New evidence includes updated rates of annual HIV infection in the United States, which are now around 40,000 incident cases per year (Centers for Disease Control and Prevention, 2021, p. 51) and updated evidence on the ability of antiretroviral therapy to reduce transmission of HIV (Rodger et al 2019). There is also new research on disparities in rates of viral suppression by race, ethnicity, and gender Buchacz et al. 2020; Buchacz et al. 2018; Geter et al. 2018). These studies show disparities in rates of viral suppression by race and ethnicity among both men who have sex with men (MSM) and women, with Black and Hispanic or Latino/a study participants having lower rates of viral suppression than White participants.

There are also updated guidelines from the Department of Health and Human Services (DHHS) Panels on Antiretroviral Guidelines for Adults and Adolescents and Children Living with HIV on defining viral suppression. The relevant content from the guidelines is included below. The guidelines also indicate the HIV viral load should be monitored at least quarterly for both adult and pediatric patients. Note: the previous viral suppression measure was chart-abstracted, while this submission is an eCQM.

Adult guidelines:

“The primary goal of antiretroviral therapy (ART) is to prevent HIV-associated morbidity and mortality. This goal is accomplished by using effective ART to achieve and maintain a plasma HIV-1 RNA (viral load) below the quantification limits of commercially available assays. Durable viral suppression improves immune function and overall quality of life, lowers the risk of both AIDS-defining and non-AIDS-defining complications, and allows persons with HIV to live a lifespan approaching that of persons without HIV.” (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2021, E-1).

“ART is recommended for all individuals with HIV to reduce the morbidity and mortality associated with HIV infection and to prevent HIV transmission to sexual partners and infants (AI). ART should be initiated as soon as possible after HIV diagnosis (AI).” (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2021, p. E-2).

“The guidelines and the AIDS Clinical Trials Group (ACTG) now define virologic failure as a confirmed viral load >200 copies/mL- a threshold that eliminates most cases of apparent viremia caused by viral load blips or assay variability.” (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2021, P. C-6).

“Individuals who are adherent to their ARV regimen and do not harbor resistance mutations to the component drugs can generally achieve suppression 8 to 24 weeks after ART initiation; rarely, in some patients it may take longer.” (Panel on Antiretroviral Guidelines for Adults and Adolescents, 2021, C-6).

Pediatric guidelines:

“Based on accumulated experience with currently available assays, the current definition of virologic suppression is a plasma viral load below the detection limit of the assay used (generally <20 to 75 copies/mL).” (Panel on Antiretroviral Therapy and Medical Management of Children Living with HIV, 2020, p. D-5).

“The Panel on Antiretroviral Therapy and Medical Management of Children Living with HIV finds value in continuing to perform viral load testing every 3 to 4 months to provide enhanced monitoring of adherence or disease progression among children and adolescents.” (D-3 of guideline)

CITATIONS:

Buchacz, K., Armon, C., Palella, F. J., Novak, R. M., Fuhrer, J., Tedaldi, E., . . . Investigators, H. O. S. H. (2020). The HIV Outpatient Study-25 Years of HIV Patient Care and Epidemiologic Research. *Open Forum Infect Dis*, 7(5), ofaa123. <https://doi.org/10.1093/ofid/ofaa123>

Buchacz, K., Armon, C., Tedaldi, E., Palella, F. J., Novak, R. M., Ward, D., . . . Investigators, H. O. S. (2018). Disparities in HIV Viral Load Suppression by Race/Ethnicity Among Men Who Have Sex with Men in the HIV Outpatient Study. *AIDS Res Hum Retroviruses*, 34(4), 357-364. <https://doi.org/10.1089/AID.2017.0162>

Centers for Disease Control and Prevention (2021). HIV Surveillance Report, 2019. <http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html>

Geter, A., Sutton, M. Y., Armon, C., Durham, M. D., Palella, F. J., Tedaldi, E., . . . Investigators, H. O. S. (2018). Trends of racial and ethnic disparities in virologic suppression among women in the HIV Outpatient Study, USA, 2010-2015. *PLoS One*, 13(1), e0189973. <https://doi.org/10.1371/journal.pone.0189973>

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

Panel on Antiretroviral Therapy and Medical Management of Children Living with HIV. Guidelines for the Use of Antiretroviral Agents in Pediatric HIV Infection, 2022. Available at <https://clinicalinfo.hiv.gov/sites/default/files/guidelines/documents/pediatric-arv/guidelines-pediatric-arv.pdf>. Updated October 11, 2022.

Rodger, A. J., Cambiano, V., Bruun, T., Vernazza, P., Collins, S., Degen, O., . . . Group, P. S. (2019). Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. *Lancet*, 393(10189), 2428-2438. [https://doi.org/10.1016/S0140-6736\(19\)30418-0](https://doi.org/10.1016/S0140-6736(19)30418-0)

[Response Ends]

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.

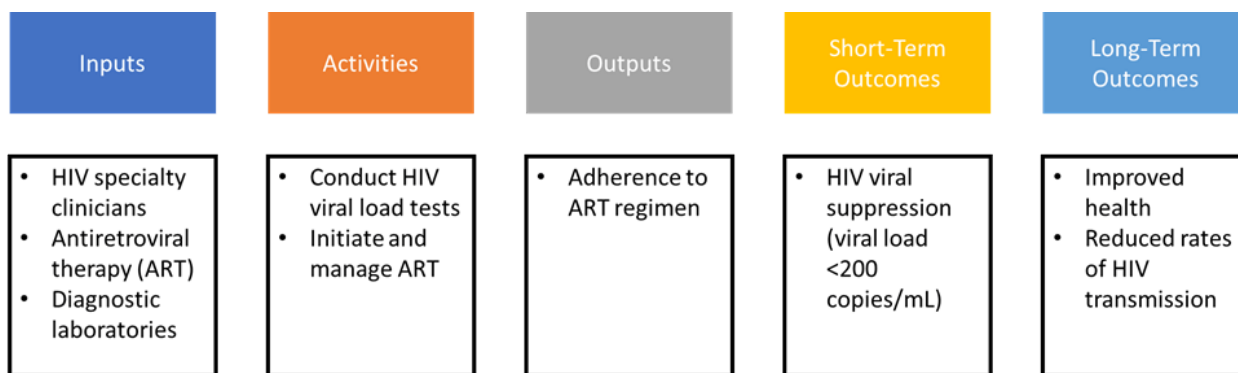
1a.01. Provide a logic model.

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]

Current Submission:

Figure 1a.01 Logic Model



The HIV "continuum of care" is the process of HIV testing, linkage to HIV care, initiation of ART, adherence to treatment, retention in care, and virologic suppression (Gardner et al 2011). Inputs to the process include HIV specialty clinicians, antiretroviral therapy (ART), and diagnostic laboratories. These inputs feed into the following activities: HIV specialty clinicians refer their patients to diagnostic laboratories, which conduct HIV viral load tests; and HIV specialty clinicians initiate and manage ART. These activities result in the output of patient adherence to their ART regimen (Meyers et al., 2019). This output results in the short-term outcome of HIV viral suppression, defined as a viral load < 200 copies/mL (Byrd et al., 2019). This short-term outcome leads to the longer-term outcomes of improved health and reduced rates of HIV transmission (Cohen et al., 2011).

ALT-TEXT:

Figure 1a.01 shows the inputs, activities, outputs, short-term outcomes, and long-term outcomes involved in the "HIV continuum of care" that result in HIV viral suppression. The inputs include HIV specialty clinicians, antiretroviral therapy (ART), and diagnostic laboratories. The activities include conducting HIV viral load tests and initiating and managing ART, and the output is adherence to ART. The short-term outcomes are HIV viral suppression and the long-term outcomes are improved health and reduced rates of HIV transmission.

CITATIONS:

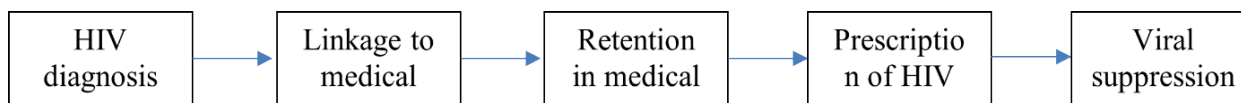
Byrd, K. K., Hou, J. G., Hazen, R., Kirkham, H., Suzuki, S., Clay, P. G., Bush, T., Camp, N. M., Weidle, P. J., Delpino, A., & Patient-Centered HIV Care Model Team (2019). Antiretroviral Adherence Level Necessary for HIV Viral Suppression Using Real-World Data. *Journal of acquired immune deficiency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

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

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

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

Previous Submission:



Although the above diagram outlines the sequential steps of medical care that people living with HIV go through from initial diagnosis to achieving the goal of viral suppression. For some patients, this is a linear path with sustained viral suppression for many years. For other patients, there may be years between diagnosis and linkage.

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

[Response Ends]

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

Describe how and from whom input was obtained.

[Response Begins]

Current Submission:

HIV viral suppression is a long-standing priority outcome among the HIV community in the United States and around the world. The National HIV/AIDS Strategy for the United States from 2022-2025, developed by the White House Office of National AIDS Policy with input from the HIV community across the United States, prioritizes increasing HIV viral suppression rates to 95%. This goal builds on the goal that was set forth by the United Nations AIDS Programme in 2014, in coordination with stakeholders, to achieve at least 90% suppression among all people receiving antiretroviral therapy worldwide (UNAIDS 2014).

CITATIONS:

The White House. (2021). National HIV/AIDS Strategy for the United States 2022–2025. <https://files.hiv.gov/s3fs-public/NHAS-2022-2025.pdf>

UNAIDS. Geneva: UNAIDS. (2014). 90-90-90: an ambitious treatment target to help end the AIDS epidemic. https://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf

[Response Ends]

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

[Response Begins]

Current Submission:

Antiretroviral therapy (ART) reduces HIV-associated morbidity and mortality by maximally inhibiting HIV replication, as defined by achieving and maintaining plasma HIV RNA (viral load) below levels detectable by commercially available assays (Hogg et al 2001; Lundgren et al., 2015). ART has also been shown to reduce transmission of HIV (Rodger et al 2019). Prior analyses have shown that provider can explain a significant amount of variation in viral suppression among Medicaid enrollees (Meyers et al 2019). Further, interventions such as providers asking for self-reported ART adherence, the use of once-daily ART regimens, reminder devices, and education and counseling are all recommended strategies for improving ART adherence (Thompson et al. 2012).

CITATIONS:

Hogg, R. S., Yip, B., Chan, K. J., Wood, E., Craib, K. J., O'Shaughnessy, M. V., & Montaner, J. S. (2001). Rates of disease progression by baseline CD4 cell count and viral load after initiating triple-drug therapy. JAMA, 286(20), 2568-2577. <https://doi.org/10.1001/jama.286.20.2568>

Lundgren, J. D., Babiker, A. G., Gordin, F., Emery, S., Grund, B., Sharma, S., . . . Group, I. S. S. (2015). Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection. *N Engl J Med*, 373(9), 795-807. <https://doi.org/10.1056/NEJMoa1506816>

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

Rodger, A. J., Cambiano, V., Bruun, T., Vernazza, P., Collins, S., Degen, O., . . . Group, P. S. (2019). Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. *Lancet*, 393(10189), 2428-2438. [https://doi.org/10.1016/S0140-6736\(19\)30418-0](https://doi.org/10.1016/S0140-6736(19)30418-0)

Thompson, M. A., Mugavero, M. J., Amico, K. R., Cargill, V. A., Chang, L. W., Gross, R., Orrell, C., Altice, F. L., Bangsberg, D. R., Bartlett, J. G., Beckwith, C. G., Dowshen, N., Gordon, C. M., Horn, T., Kumar, P., Scott, J. D., Stirratt, M. J., Remien, R. H., Simoni, J. M., & Nachega, J. B. (2012). Guidelines for Improving Entry Into and Retention in Care and Antiretroviral Adherence for Persons With HIV: Evidence-Based Recommendations From an International Association of Physicians in AIDS Care Panel. *Annals of Internal Medicine*, 156(11), 817–833. <https://doi.org/10.7326/0003-4819-156-11-201206050-00419>

Previous Submission:

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

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

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

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

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

[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]

Current Submission:

HIV is a communicable infection that leads to a progressive disease with a long asymptomatic period. Approximately 40,000 persons in the United States are newly infected with HIV each year (Centers for Disease Control and Prevention, 2021, p. 51). Without treatment, most persons develop acquired immunodeficiency syndrome (AIDS) within 10 years of HIV infection.

HIV viral suppression is a long-standing priority outcome among the HIV community in the United States and around the world. The National HIV/AIDS Strategy for the United States from 2022-2025, developed by the White House Office of National AIDS Policy with input from the HIV community across the United States, prioritizes increasing HIV viral suppression rates to 95% (The White House 2020). The DHHS Panel on Antiretroviral Guidelines for Adults and Adolescents defines viral suppression as a viral load below the lower limits of detection in its guidelines on virologic failure, and it defines viral suppression as a viral load of less than 200 copies/mL as part of its guidelines for the use of antiretroviral therapy to prevent HIV transmission (Panel on Antiretroviral Guidelines for Adults and Adolescents 2022).

Antiretroviral therapy (ART) delays the progression to AIDS and increases the length of survival. ART reduces HIV-associated morbidity and mortality by maximally inhibiting HIV replication to achieve viral suppression (Hogg et al 2001; Lundgren et al., 2015). ART has also been shown to reduce transmission of HIV (Rodger et al 2019). Studies show disparities in rates of viral suppression by race and ethnicity among MSM and among women, with Black and Hispanic or Latino/a study participants having lower rates of viral suppression than White participants (Buchacz et al. 2020; Buchacz et al. 2018; Geter et al. 2018). This measure will help providers direct their attention and quality improvement efforts towards improving HIV viral suppression rates.

CITATIONS:

Buchacz, K., Armon, C., Palella, F. J., Novak, R. M., Fuhrer, J., Tedaldi, E., . . . Investigators, H. O. S. H. (2020). *The HIV Outpatient Study-25 Years of HIV Patient Care and Epidemiologic Research*. *Open Forum Infect Dis*, 7(5), ofaa123. <https://doi.org/10.1093/ofid/ofaa123>

Buchacz, K., Armon, C., Tedaldi, E., Palella, F. J., Novak, R. M., Ward, D., . . . Investigators, H. O. S. (2018). *Disparities in HIV Viral Load Suppression by Race/Ethnicity Among Men Who Have Sex with Men in the HIV Outpatient Study*. *AIDS Res Hum Retroviruses*, 34(4), 357-364. <https://doi.org/10.1089/AID.2017.0162>

Centers for Disease Control and Prevention (2021). *HIV Surveillance Report, 2019*. <http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html>

Geter, A., Sutton, M. Y., Armon, C., Durham, M. D., Palella, F. J., Tedaldi, E., . . . Investigators, H. O. S. (2018). *Trends of racial and ethnic disparities in virologic suppression among women in the HIV Outpatient Study, USA, 2010-2015*. *PLoS One*, 13(1), e0189973. <https://doi.org/10.1371/journal.pone.0189973>

Hogg, R. S., Yip, B., Chan, K. J., Wood, E., Craib, K. J., O'Shaughnessy, M. V., & Montaner, J. S. (2001). *Rates of disease progression by baseline CD4 cell count and viral load after initiating triple-drug therapy*. *JAMA*, 286(20), 2568-2577. <https://doi.org/10.1001/jama.286.20.2568>

Lundgren, J. D., Babiker, A. G., Gordin, F., Emery, S., Grund, B., Sharma, S., . . . Group, I. S. S. (2015a). *Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection*. *N Engl J Med*, 373(9), 795-807. <https://doi.org/10.1056/NEJMoa1506816>

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

Rodger, A. J., Cambiano, V., Bruun, T., Vernazza, P., Collins, S., Degen, O., . . . Group, P. S. (2019). *Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study*. *Lancet*, 393(10189), 2428-2438. [https://doi.org/10.1016/S0140-6736\(19\)30418-0](https://doi.org/10.1016/S0140-6736(19)30418-0)

The White House. (2021). National HIV/AIDS Strategy for the United States 2022–2025. <https://files.hiv.gov/s3fs-public/NHAS-2022-2025.pdf>

Previous Submission:

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

Antiretroviral therapy reduces HIV-associated morbidity and mortality by maximally inhibiting HIV replication (as defined by achieving and maintaining plasma HIV RNA (viral load) below levels detectable by commercially available assays). Viral suppression is a main goal of HIV treatment and an indicator of treatment success and reduction of potential HIV transmission. It is directly related to:

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

[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 187 unique clinicians and 3,056 patients, of which 47 clinicians had at least 11 patients—the minimum sample size requirement outlined in the CMS cell suppression policy. The data for these 47 clinicians included 2,995 unique patients. We provide the distribution of the measure performance scores for all clinicians and clinicians with at least 11 patients eligible for the denominator below. In the remainder of this document, we will focus on clinicians with at least 11 patients.

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

Sample	Clinicians	Patients	Mean	SD	Min	p10	p20	p30	p40	p50	p60	p70	p80	p90	Max	IQR	
All clinicians	187	3,056	72.3	33.3	0.0	0.0	50	66.1	77.5	85.2	93.2	100	100	100	100	50	

Sample	Clinicians	Patients	Mean	SD	Min	p10	p20	p30	p40	p50	p60	p70	p80	p90	Max	IQR	
Clinicians with 11+ patients	47	2,995	85.2	10.9	46.2	70	79.3	81.7	83.9	87.9	90.8	91.5	94.5	95.2	100	12.4	

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

These data reflect Ryan White HIV/AIDS Program patients, and thus reflect a sample of patients with higher rates of viral suppression than the national population of people with HIV (HRSA 2022, CDC 2020).

ALT-TEXT:

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

CITATIONS:

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

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

[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.) A total of 47 clinicians and 2,995 unique patients from 7 test sites are included in the analysis below. During testing, measure performance was stratified to assess whether there were disparities in viral suppression by patients' age (< 50 years vs. ≥ 50 years), HIV transmission category (men who have sex with men [MSM], injection drug use [IDU], which includes both IDU and MSM and IDU, and Other transmission), race (White vs. Black), and ethnicity (Hispanic or Latino vs. not Hispanic or Latino). Table 1b.04 summarizes the results of the analysis.

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

Patient Group	Clinicians	Patients	Mean	Standard deviation	Min	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	Max	IQR
AGE<50	47	1,472	82.4	12.5	40.0	68.7	75.0	76.4	80.2	82.7	86.9	88.5	91.8	100.0	100.0	14.6
AGE≥50	47	1,523	87.7	13.0	50.0	67.9	81.9	84.2	89.7	90.6	94.1	95.4	100.0	100.0	100.0	13.3
MSM	41	1,218	84.7	13.2	50.0	66.7	71.4	81.0	85.7	88.9	90.0	92.9	95.8	100.0	100.0	15.1
IDU	40	234	82.2	22.6	0.0	50.0	66.7	74.1	84.1	90.1	100.0	100.0	100.0	100.0	100.0	31.4
Other transmission	41	992	87.8	13.6	50.0	66.7	78.6	85.7	90.9	93.1	93.8	95.2	100.0	100.0	100.0	11.0
Black	47	1,351	79.7	21.2	0.0	60.0	73.2	78.9	81.6	85.7	88.5	90.2	92.6	94.8	100.0	15.1
White	47	1,520	86.7	12.4	50.0	66.7	75.0	82.7	87.0	90.9	92.7	95.3	99.3	100.0	100.0	20.3
Not Hispanic or Latino	47	2,559	84.7	11.9	50.0	65.9	78.9	80.9	84.7	87.1	90.0	92.4	94.4	96.1	100.0	12.9
Hispanic or Latino	36	429	89.7	19.1	0.0	77.5	84.6	88.2	93.8	99.1	100.0	100.0	100.0	100.0	100.0	13.2

Notes: Results are for clinicians with ≥11 patients eligible for the denominator.

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

Across all clinicians, mean clinician-level performance rates varied by age, HIV transmission group, and race/ethnicity. Patients under age 50 had lower rates of viral suppression (mean=82.4%) as compared to patients age 50 and older (mean=87.7%). The differences in the measure score by patients' ages were statistically significant (p=0.05), which reflects the lower rates of HIV viral suppression among younger patients based on national surveillance data (HRSA 2022). Black patients had lower rates of viral suppression than white patients (79.7% vs. 86.7%), while non-Hispanic patients had lower rates of viral suppression than Hispanic patients (84.7% vs. 89.7%). Patients whose HIV transmission group is IDU had lower rates of suppression (82.2%) than those whose transmission group was MSM (84.7%) or other (87.8%). However, the differences in the measure scores by ethnicity (p=0.18), race (p=0.06), and HIV transmission group (p=0.42) did not reach statistical significance at the p≤0.05 level. We observed more variability in the clinician-level measure rates for Black patients (mean=79.7, SD=21.2) and IDU patients (mean=82.2, SD=22.6) relative to other strata. Larger variation in the measure rates for IDU patients can be attributed to the relatively small sample for that stratum, whereas variability in the measure rates for Black patients may reflect differences in care within this group of patients. Please refer to the 2b.02 for the interpretation.

ALT-TEXT:

Table 1b.04 describes the measure performance rate age, virus transmission model, race, and ethnicity.

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

CITATIONS:

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

[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.

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

[Response Begins]

Yes

[Yes Please Explain]

Denominator criteria require the eligible encounter to occur within the first 8 months of the measurement period to allow enough time for the clinician to help the patient to achieve viral suppression before the end of the measurement period. The previous denominator did not limit the eligible encounter to the first 8 months. This allows for time for patients who 1) change providers to re-establish viral suppression and obtain viral load test, which is recommended quarterly; or 2) are newly diagnosed to take HIV medications long enough to achieve viral suppression. Denominator criteria also require an HIV diagnosis to occur within or prior to the first 3 months of the measurement period to allow enough time for newly diagnosed, ART-naïve patients to achieve initial suppression, which can take up to 26 weeks (Saag et al., 2020). The previous denominator did not limit the timing of the HIV diagnosis.

CITATIONS:

Saag, M. S., Gandhi, R. T., Hoy, J. F., Landovitz, R. J., Thompson, M. A., Sax, P. E., . . . Volberding, P. A. (2020). Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults: 2020 Recommendations of the International Antiviral Society-USA Panel. JAMA, 324(16), 1651-1669. <https://doi.org/10.1001/jama.2020.17025>

[Response Ends]

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

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

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

[Response Begins]

Not applicable

[Response Ends]

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]

HIV viral suppression

[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, regardless of age, diagnosed with HIV prior to or during the first three months of the measurement period, with an eligible encounter in the first eight months of the measurement period, who have a last HIV viral load test has result of less than 200 copies/mL during 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]

Health and Functional Status: Change

Primary Prevention

[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]

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 eQCM 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: 3210e_CMS314-v0-1-035-QDM-5-6.zip

sp.11. Attach the simulated testing attachment.

All eQCMs 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: 3210e_CMS314v0.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: 3210e_CMS314+HIV+Viral+Loa+(2022-10-31+16-11-09).xlsx

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

sp.13. State the numerator.

Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).

DO NOT include the rationale for the measure.

[Response Begins]

Patients with a last HIV viral load test result of less than 200 copies/ml during the measurement period

[Response Ends]

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

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

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

Codes for qualifying viral load tests are in the attached file (see also value sets in sp.12 and specifications in sp.10).
Measurement period is equivalent to a calendar year.

File is CMS314+HIV+Viral+Loa+(2022-10-31+16-11-09).xlsx (attached)

[Response Ends]

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

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

All patients, regardless of age, diagnosed with HIV prior to or during the first three months of the measurement period with at least one eligible encounter in the first eight months of the measurement period.

[Response Ends]

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

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

All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

Codes identifying qualifying HIV diagnoses and eligible encounter codes are in the attached file (see also value sets in sp.12 and specifications in sp.10).

Patient age

HIV diagnosis date

File is CMS314+HIV+Viral+Loa+(2022-10-31+16-11-09).xlsx

[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 in the first eight months of the measurement period
2. Retain all patients diagnosed with HIV during the first three months of the measurement period or any time prior.
3. Patients meeting these criteria are in the denominator.

Numerator

1. Identify denominator eligible patients with an HIV viral load test during the measurement period.
2. Identify the last HIV viral load test during the measurement period
3. If the last HIV viral load test value is less than 200 copies/mL and/or below the lower limit of detection, the patient is included in the numerator. If the last HIV viral load test value is greater than or equal to 200 copies/mL, 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 47 clinicians and 2,995 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 47 clinicians. As a part of reliability testing in order to add rigor to the limited number of unique clinicians, we also conducted bootstrap resampling. This approach was used to test the stability of the measure rates over 2,000 replications of the initial sample. Results from the bootstrap testing (section 2a.11) support the generalizability of the findings.

1. **Geographic region and urban/rural communities served**
 1. Four Northeast; two Midwest; one South
 2. Five urban; two combined urban and rural
2. **Clinic types**
 1. Two hospital or university-based clinics
 2. Four publicly funded community health centers
 3. One other community-based service organization
3. **Electronic health record (EHR)**
 1. eClinicalWorks (3)
 2. EPIC (2)
 3. NextGen (1)
 4. 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 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]

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

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

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

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

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

[Response Begins]

No

[Response Ends]

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

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

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

[Response Begins]

No additional risk adjustment analysis included

[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]

Current Submission:

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

Previous Submission:

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

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

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

[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]

Current Submission:

01-01-2021 – 12-31-2021

Previous Submission:

The Bonnie test environment simulates the year 2012 as the measurement period.

[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]**Current Submission:**

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 47 clinicians were included in testing. These 47 clinicians had a total of 2,995 patients included in the measure denominator. Table 2a.05 breaks down the characteristics of the participating sites included in the beta testing of the measure.

Table 2a.05. Test site characteristics

Site	Provider type	Region	EHR	# of clinicians with 11+ patients	# of patients
Site 1	Publicly funded community health center	NE	eClinical Works	5	136
Site 2	Publicly funded community health center	NE	eClinical Works	3	157
Site 3	Hospital or university-based clinic	NE	EPIC	21	592
Site 4	Publicly funded community health center	SO	OCHIN EPIC	6	516
Site 5	Other community-based service organization	NE	NextGen	3	60
Site 6	Publicly funded community health center	MW	Athena Health	3	484
Site 7	Hospital or university-based clinic	MW	eClinical Works	6	1050

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

ALT-TEXT:

Table 2a.05 provides characteristics of the seven test sites, including provider type, provider region, provider electronic health record system, as well as the number of clinicians (for clinicians with at least 11 patients) and patients in the sample the measure developer received from each provider.

Previous Submission:

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

[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]

The full analytic sample extracted from the EHR included 3,056 patients attributed to 187 clinicians within the measurement period from 7 sites that are Ryan White HIV/AIDS Program participants. This sample represents all patients with HIV that had any encounter during the measurement period. The measure is specified to require a minimum denominator of 11 patients during the measurement period, to follow the CMS cell size suppression policy stating that no cell can be reported that allows a value of 1 to 10 reported. The restricted sample, used for testing, includes 47 clinicians (25.1% of the initial number of clinicians) and 2,995 patients (98.0% 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 HIV Viral Suppression measure rate of 85.2%.

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

1. 3% (88) patients were under the age of 25, and 97% (2,910) patients were over the age of 25.
2. 49% (1,472) patients were under the age of 50, and 51% (1,523) patients were over the age of 50.
3. Broken out by HIV transmission group, 41% (1,218) of patients' transmission group was men who have sex with men (MSM), 6% (186) 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 35% (1,015) of patients' transmission group was other, while 18% (551) of patients were missing information on HIV transmission group.
4. 25% (748) patients were cisgender women, 74% (2,232) patients were cisgender men, and 0.5% (15) patients were transgender women.

Previous Submission:

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

1. Patient name
2. Date of birth
3. Race
4. Ethnicity
5. Gender
6. Payer

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

1. Diagnosis
2. Laboratory tests and associated results
3. Encounters

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

The breakdown of test bundle demographics for the 34 patients included (represented by number of patients/percentage of bundle): males 23/68%; females 11/32%; American Indian/Alaska Native 1/3%; Asian 1/3%; Black/African American 15/44%; Native Hawaiian/Pacific Islander 0/0%; White 9/26%; Hispanic/Latino 8/24%; younger than 13 1/3%; 13-17 years old 1/3%; 18-24 years old 2/6%; 25-34 years old 6/18%; 35-44 years old 6/18%; 45-54 years old 10/29%; 55-65 years old 6/18%; older than 65 2/6%.

Full details on the Bonnie synthetic patient bundle used to test this measure, including human-readable and QRDA Category 1 format documents for each synthetic patient record, are included in the Bonnie testing attachment.

[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]

Current Submission:

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

Reliability: To assess reliability, we used the EHR data from each of seven sites covering the period between January 1, 2021, and December 31, 2021. For the reliability analysis (and all other clinician-level testing) we restricted the sample to clinicians who saw at least 11 patients during the measurement period.

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

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

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

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

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

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

Previous Submission:

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

[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]

Current Submission:

We collected information on the following variables using data extracted from hospital EHR systems: age, sex, race, ethnicity, payer, and HIV transmission category. We examined disparities in the measure rates by age, race, ethnicity, and HIV transmission category. Section 1b.4 describes those results.

Previous Submission:

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

[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]

Current Submission:

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

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

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

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

We used split-half and test-retest approaches to examine stability of the measure scores within the same clinicians. The split-half method involves comparing the measure scores for two independent samples of patients within the same provider. For this method, we randomly split the sample of patients within each clinician into two mutually exclusive samples with equal or nearly equal size, resulting in two samples that cover the same one-year period but with case volume the size of a measure that would be calculated with six months of data. **Thus, each clinician appeared in the sample twice, but with an entirely different set of patients.** Then, we estimated Spearman- and intraclass correlations between the measure rates within two samples. We also computed the Spearman-Brown correction to account for the attenuation of the Spearman correlation due to dividing the original sample of patients in two halves.

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

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

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

CITATIONS:

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

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

Previous Submission:

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

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

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

[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]

Current Submission:

Table 2a.11-A summarizes the mean and range of the signal-to-noise reliability statistics for the HIV Viral Suppression measure, which was calculated separately for each clinician. The mean signal-to-noise reliability across all 47 clinicians with at least 11 patients eligible for the denominator was 0.90, with the standard deviation of 0.10, suggesting that the measure is highly reliable.

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	47	0.628	0.90	0.10	0.70	0.75	0.85	0.94	0.98	0.99	1.00	1.00

Notes: Results are provided for clinicians with ≥11 patients eligible for the denominator (n = 47). Min=minimum, SD=standard deviation, Max=maximum

ALT-TEXT:

Table 21.11-A provides results of the signal-to-noise reliability testing for clinicians with at least 11 patients eligible for the denominator. The table shows the number of clinicians, and the mean, standard deviation, minimum and maximum scores on the measure, as well as the measure scores for the 5th, 10th, 25th, 50th, 75th, 90th and 95th

percentiles. Mean reliability in a sample was very high and measure scores for most clinicians in a sample were highly reliable.

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

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

Sample	Spearman rank-order correlation	Spearman-Brown correction for the sample size attenuation	Intra-class correlation
Clinicians >= 11 patients	0.967	0.983	0.932

Notes: Results are provided for clinicians with ≥ 11 patients eligible for the denominator (n = 47)

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 HIV Viral Suppression measure, which were calculated using the bootstrap method in 1,000 pairs of bootstrap samples. In this analysis, the mean reliability statistics exceeded the 0.9 threshold, indicating very high stability of the measure scores across 1,000 pairs of patient samples.

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	0.973	0.897	0.951	0.959	0.967	0.975	0.981	0.985	0.987	0.993
Clinicians >= 11 patients	Spearman-Brown correction	0.986	0.945	0.975	0.979	0.983	0.987	0.990	0.993	0.993	0.997
Clinicians >= 11 patients	intra-class correlation	0.944	0.811	0.899	0.914	0.933	0.948	0.961	0.970	0.973	0.987

Notes: Results are provided for clinicians with ≥ 11 patients eligible for the denominator (n = 47)

Our findings regarding the minimum sample size for the measure indicate the median sample size per clinician necessary to reach signal-to-noise reliability of 0.7 in our sample was 7 patients. With a sample size of 10 patients, 75 percent of clinicians would reach the 0.7 threshold for signal-to-noise reliability. Therefore, assuming our findings are generalizable to the universe of reporting clinicians, our findings indicate that using the CMS cell suppression policy to set a minimum sample size of 11 patients would allow us to produce statistically valid comparisons between clinicians.

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.

Previous Submission:

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

Overall reliability scores by year, 2010-2014

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

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

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

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

[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]

Current Submission:

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

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

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

indicate excellent reliability. The ICCs exceeding 0.9 obtained from the split-half and test-retest testing methods indicate excellent reliability of the measures.

Finally, according to Cohen's (Cohen, 1992) effect-size criteria, the Spearman correlations above 0.8 indicate a large effect size, thus also supporting the claim that the HIV Viral Load Suppression measure is highly reliable.

CITATIONS:

Adams JL. *The Reliability of Provider Profiling: A Tutorial*. Santa Monica, CA: RAND Corporation; 2009.

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

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

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

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

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

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

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

Previous Submission:

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

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

[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]

Current Submission:

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 agreement for measure data elements 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 (considered the 'gold-standard')

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

We calculated the raw agreement (percentage agreement) and the chance-corrected agreement (Gwet's AC1) between the two data sources. The interpretation of the AC1 statistic is the same as that of Cohen's Kappa, but AC1 is a more robust measure of interrater reliability. Kappa is sensitive to classification probabilities which in some cases lead to the low chance-corrected agreement despite the high observed agreement (the so-called Kappa paradox). This situation does not occur when using AC1 (Quarfoot and Levine 2016). Higher values for agreement statistics demonstrate that the structured EHR data used to calculate the measure have accuracy similar to looking at the medical record overall, including clinical notes, documents, and other fields that convey information about the patient but cannot be used to calculate 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 a formal poll during a meeting of the Technical Expert Panel (TEP) that was convened during the development of the HIV Viral Suppression 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 HIV viral suppression: transmission category (injection drug use [IDU] versus non-IDU) 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>

Previous Submission:

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

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

1. **Clinical relevance.** References cited within the chart abstracted measure specification were used to design clinically relevant, realistic patient profiles for the measure's target population. This approach ensured the eCQM logic maintained alignment with the clinical intent of the chart abstracted measure.
2. **100% logic coverage:** The resulting bundle of synthetic patients collectively includes all data elements and conditions logic that are specified within the measure logic, including at least one patient evaluating against each measure population pathway. Fully testing the measure logic increases test rigor and mitigates risk of unexpected outcomes.
3. **Edge case testing.** Edge cases refer to those data elements that test the upper or lower boundary of measure logic conditions, e.g. a diagnosis starting on the latest qualifying date or an HIV viral load result equal to the highest qualifying value. Edge cases are designed to test each edge that exists within each measure population.
4. **Negative testing.** Negative testing involves use of test cases do not evaluate positively against measure logic, but are otherwise clinically relevant and realistic, e.g. scenarios where an HIV diagnosis was not documented or an HIV viral load was performed without a documented result. Negative testing further validates measure logic by accurately evaluating patients against expected outcomes and simulating the effect of missing data on measure results.

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

[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 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

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

ALT-TEXT:

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

Face validity. We found that six of the seven clinicians interviewed (86%) agreed that the measure can distinguish good from poor quality of care.

Construct validity. Among patients attributed to clinicians with at least 11 patients, the mean rate of viral suppression was 82.4% for those under 50 as compared to 87.7% for those 50 years or older with an effect size of 0.415 using Cohen's D, indicating a moderate effect. The mean rate of viral suppression was 81.2% for those whose HIV transmission group was IDU as compared to 88.1% for those whose HIV transmission group was not IDU, with an effect size of 0.404 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 Viral Suppression	S.D.	Cohen's D	t-test p-value
Age < 50 years	82.4	12.3	0.415	0.047
Age >= 50 years	87.7	13.0		
Non-IDU HIV transmission category	88.1	8.5	0.404	0.085
IDU HIV transmission category	81.2	22.8		

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 viral suppression 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.

Previous Submission:

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

[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]

Current Submission:

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). [add summary sentence or two stating the measure showed good evidence of data element validity and known group validity]

Face validity. Most respondents strongly agreed or agreed that the measure score is an accurate reflection of quality (4 out of 7 respondents) and that the measure score can be used to distinguish between good and poor quality of care (6 out of 7 respondents). Further, among the technical expert panel (TEP) convened for the development of this measure, 100% agreed that the measure was important and related to quality of care. These results demonstrate high face validity of the measure.

Data element validity. The AC1 values calculated through data element validity testing suggest high levels of agreement between the data extract generated from the EHR systems and the manually abstracted data. We observed the 96.4 percent average agreement across all data elements or higher for all data elements. The average chance-corrected agreement captured by the AC1 statistic was 95.2 percent. These statistics indicate very high data element validity. Observed agreement was very high (above 98 percent) for all data elements with the exception for the HIV diagnosis date, for which the observed agreement was 75 percent. As a sensitivity test, we compared the HIV diagnosis dates in both data sources (i.e., EHR and chart data). We found that in all cases, observed differences between the HIV diagnosis dates did not affect whether or not the patient would have been included in the denominator because all of the dates occurred prior to the measurement period.

Known-group validity. The differences in the measure rates between patient subgroups by age and HIV transmission category are consistent with the observed literature. We found that younger patients (age < 50 years) had had viral suppression rates 5.3 percentage points lower than older patients (age 50+ years). This is consistent with national Ryan White HIV/AIDS Program data from 2021 showing that individuals with HIV who are ages 15-19 and 20-24 have HIV viral suppression rates of 81.7 and 82.7%, respectively, which is lower than individuals with HIV at older ages (e.g., ages 50-54 and 55-59 have HIV viral suppression rates of 91.1% and 92.5%, respectively) [Health Resources and Services Administration, 2021].

We also found that patients whose HIV transmission category was Injection Drug Use had viral suppression rates approximately 7 percentage point lower than patients in the non-IDU transmission category. This is also consistent with the literature indicating that viral suppression rates were about 6% lower in patients whose HIV transmission group was Injection Drug Use (IDU) using the most recent HIV viral load measure using data from eight HIV clinical cohorts across 1997 to 2015 (Nance et al, 2018). Thus, observed differences in the viral suppression rates observed in our sample are in line with the literature, providing evidence of measure validity.

CITATIONS:

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

*Nance, R. M., Delaney, J. A. C., Simoni, J. M., Wilson, I. B., Mayer, K. H., Whitney, B. M., . . . Crane, H. M. (2018). HIV Viral Suppression Trends Over Time Among HIV-Infected Patients Receiving Care in the United States, 1997 to 2015: A Cohort Study. *Ann Intern Med*, 169(6), 376-384. <https://doi.org/10.7326/M17-2242>*

Previous Submission:

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

[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]

Current Submission:

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

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

CITATIONS:

Centers for Medicare & Medicaid Services. Analysis of Topped-Out Measures Finalized for the PY 2016 ESRD QIP. (2014). Updated June 19, 2014. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ESRDQIP/Downloads/AnalysisofTopped-OutMeasuresFinalizedforthePY2016ESRDQIP.pdf>.

Previous Submission:

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

[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]

Current Submission:

Based on the sample of 47 clinicians with at least 11 patients eligible for denominator, the HIV Viral Load Suppression measure rates in our sample ranged from 46.2% to 100.0% (with a median of 87.9% and an interquartile range of 80.6% and 93.0%). Thus, there is variation in measure scores across clinicians.

Table 2b.06-A. Performance distribution of the HIV Viral Suppression measure rates

Sample	Clinicians	Patients	Mean	SD	Min	p10	p25	p50	p75	p90	Max	IQR
Clinicians with 11+ patients	47	2,995	85.2	10.9	46.2	70.0	80.6	87.9	93.0	95.2	100.0	12.4

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 47 clinicians, 13 (28% of all clinicians in the sample) were statistically better, and 2 clinicians (4.3%) were worse than the sample average, which is conceptually equivalent to an “average-performing clinician” in a sample. Distribution of the performance categories shown in Table 2b.06-B suggests that improvement in the measure scores is possible for nearly three quarters of clinicians whose performance scores were either no different from the sample average (68.1%) or worse than the sample average (4.3%).

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

Performance group	N and % of clinicians	Mean performance rate	
Better than the sample average	13 (27.7%)	94.9%	
No different than the sample average	32 (68.1%)	83.0%	
Worse than the sample average	2 (4.3%)	56.9%	
All Clinicians	47 (100.0%)	85.2%	

ALT-TEXT:

Table 2b.06-B shows performance distribution of the HIV viral suppression measure scores relative to the sample average for clinicians with at least 11 patients eligible for the denominator. Out of 47 clinicians in the sample, 32 clinicians had measure scores that were not significantly different from the sample average and 13 and 2 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 has a truncated coefficient of variation (TCV) equal to 0.09, which meets criterion 2 (TCV<0.10), but does not meet criterion 1 (75th percentile within 2 standard errors of the 90th percentile), and thus the measure is not considered topped out.

Table 2b.06-C. Topped out analysis of the HIV Viral Suppression measure rates.

Measure	75 th pctl.	90 th pctl.	90 th – 75 th pctl.	2x S.D. of 90 th pctl.	Criterion 1 met?	TCV	Criterion 2 met?
HIV Viral Suppression	0.93	0.95	0.02	0.01	No	0.09	Yes

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 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.

Previous Submission:

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

[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]**Current Submission:**

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 about 25 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. Further, these data reflect Ryan White HIV/AIDS Program patients, and thus reflect a sample of patients with higher rates of viral suppression than the national population of people with HIV (HRSA 2022, CDC 2020). It is likely that when this measure is applied to a broader population of patients with HIV, the performance scores are unlikely to be as high.

CITATIONS:

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

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

Previous Submission:

The table above demonstrates meaningful variability across providers, allowing for the identification of meaningful differences across sites. Specifically, the measure is able to detect providers with better or worse than median performance scores. In 2014, the bottom 10% of providers had viral suppression rates of 65.0% or lower; the top 90% of providers had viral suppression rates of 93.1% or higher. While this gap appears to be narrowing over time, a meaningful difference of 28.1 percentage points remains, demonstrating the value of the measure in identifying sites based on poor performance relative to the top performers.

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

[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]

Current Submission:

Missing data are not expected to be a threat to validity for the measure. 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 data element. For example, if a patient does not have a lab visit, we interpret this to mean that the patient did not have an eligible lab visit, rather than that the information for that visit was missing. Encounter type and dates are also required for the measure calculation. Results on missing data elements used in testing are presented below in Section 2b.09.

Previous Submission:

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

[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]

Current Submission:

As noted in 2b.08, we do not expect missing data to be a threat to validity. No data elements used in calculating the measure scores had substantial rates of missing values.

0.1% (9) of the viral load tests had a performance date but missing viral load value. This represents a very small number of tests and includes instances where the test was performed but results were inconclusive.

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.

Previous Submission:

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

[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]

Current Submission:

As noted in 2b.08, we do not expect missing data to be a threat to validity. As noted in 2b.09, no data elements used in calculating the measure scores had substantial rates of missing values.

Previous Submission:

Please see response for question 2b7.1 above. (2b7.1 which is now question 2b.08)

[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 eQMs). 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]

Current Submission:

Not applicable.

Previous Submission:

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]

Current Submission:

Not applicable.

Previous Submission:

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]

Current Submission:

Not applicable.

Previous Submission:

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]

Current Submission:

Not applicable.

Previous Submission:

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]

Not applicable.

[Response Ends]

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

[Response Begins]

[Response Ends]

Attachment: 3210e_10282022_HIVViralSuppressionFeasibilityScorecard_(1).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 HIV/AIDS 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 of our test sites, we do not expect this data element to substantially affect the feasibility of this measure. Moreover, in sites where the diagnosis date is unavailable,

the date associated with the diagnosis on the problem list should be sufficient to determine whether diagnosis occurred prior to the performance year.

[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]

eCQM has been newly tested and not currently in use.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Payment Program

Quality Improvement (internal to the specific organization)

[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]

A MIPS CQM version of this measure, Quality ID 338, is currently in use in MIPS. HRSA plans to replace the MIPS CQM version with the present eCQM.

[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]

N/A. 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]

N/A. Measure has not yet been implemented.

[Response Ends]

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

[Response Begins]

N/A. Measure has not yet been implemented.

[Response Ends]

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

[Response Begins]

N/A. 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]

N/A. 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 as an eCQM in a quality improvement program. During measure testing, clinicians at all seven test sites agreed that the measure could be used to improve quality of care at their practices. Several of these clinicians noted that their practices already track which patients have not achieved viral suppression, and they use these results to focus outreach efforts on those patients to help improve medication adherence and take other steps to help them achieve viral suppression. Some of the clinicians interviewed also said that scores on this measure could motivate low-scoring clinicians to focus on improving their viral suppression rates through actions such as improving their communication with patients and making sure they were tracking their patients' viral suppression. As noted above in 1a.01, helping patients achieve viral suppression improves patient health and reduces the risk of those patients transmitting HIV to others. HRSA HAB is considering these measures for use in CMS's Quality Payment Programs (QPP), in particular the Merit-based Incentive Payment System (MIPS), which would provide a financial incentive for clinicians to ensure their patients are virally suppressed.

[Response Ends]

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

[Response Begins]

N/A. Measure has not yet been implemented.

[Response Ends]

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

[Response Begins]

N/A. 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

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]

[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]

0405 HIV/AIDS: Pneumocystis Jiroveci Pneumonia (PCP) Prophylaxis

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

2079 HIV Medical Visit Frequency

2080 Gap in HIV Medical Visits

2082 HIV Viral Suppression

2083 Prescription of HIV Antiretroviral Therapy

3211 Prescription of HIV Antiretroviral Therapy

3010 HIV Medical Visit Frequency

[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]

The denominator population for this measure differs slightly from three related measures—3209e, 3211e, and 0409—with respect to the timing of the patient’s HIV diagnosis and eligible encounter, and these differences are due to the specific timing required for measuring viral suppression. We limit the population to patients diagnosed no earlier than three months into the performance period because our expert workgroup and the clinicians we interviewed during testing agreed that this allowed sufficient time for a clinician to work with a newly diagnosed patient to achieve viral suppression. Similarly, we limited eligible encounters to those occurring within the first eight months of the measurement period to ensure that clinicians had enough time left in the year to work with new patients to achieve viral suppression. Given that neither 3209e, 3211e, nor 0409 are currently in use in MIPS as eQMs, we do not expect these differences to meaningfully affect data collection burden. The measure also differs from the denominator population for 0409 with respect to the patient’s age because viral suppression is a relevant clinical outcome for all patients with HIV, regardless of age, while 0409 focuses on older patients who may be sexually active. Again, given that 0409 is not currently in use in MIPS as an eQCM, 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]

Not applicable

[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

Matosky, Marlene, mmatosky@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

Matosky, Marlene, mmatosky@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]

2017

[Response Ends]

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

[Response Begins]

July 2017

[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]

Spring 2023 (review of NQF endorsement)

[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]

N/A

[Response Ends]