NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0033

Corresponding Measures:

Measure Title: Chlamydia Screening in Women (CHL)

Measure Steward: National Committee for Quality Assurance

Brief Description of Measure: The percentage of women 16–24 years of age who were identified as sexually active and who had at least one test for chlamydia during the measurement year.

Developer Rationale: This measure assesses the percentage of women 16-24 years of age who were identified as sexually active and who received a test for chlamydia. The improvement in quality envisioned by the use of this measure is increased identification of untreated chlamydia infections in women that can lead to serious and irreversible complications and can be unknowingly transmitted to sexual partners. Despite the availability of effective treatments, a large proportion of sexually active individuals continue to go undiagnosed due to the disease's asymptomatic nature. Early detection, screening, and treatment have proven to be effective in managing and preventing chlamydia.

Centers for Disease Control and Prevention (CDC). 2021. "Sexually Transmitted Diseases: Chlamydia—CDC Fact Sheet." http://www.cdc.gov/std/chlamydia/STDFact-chlamydia-detailed.htm

Numerator Statement: Women who were tested for chlamydia during the measurement year.

Denominator Statement: Women 16-24 years of age who had a claim or encounter indicating sexual activity.

Denominator Exclusions: Women who received a pregnancy test to determine contraindications for medication (isotretinoin) or x-ray.

Women who were in hospice or using hospice services during the measurement year.

Measure Type: Process

Data Source: Claims, Enrollment Data

Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date**: Oct 25, 2016

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

1a. Evidence. The evidence requirements for a *structure, process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

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

Evidence Summary or **Summary of prior review in [2016]**

- In the previous submission, the developer provided updated US Preventative Services Task Force (USPSTF) (2014) recommendations for screening for chlamydia in sexually active females aged 24 years or younger and in older women who are at increased risk for infection. Evidence synthesis concluded, "Chlamydia screening in young women may reduce pelvic inflammatory disease." USPSTF notes "the studies it reviewed on the direct effects of screening for chlamydia, including one new good-quality RCT, showed mixed results. This led to the change in grade for screening for chlamydia, which is now based on "moderate" certainty of a moderate net benefit rather than "high certainty" of a substantial net benefit."
- During the discussion, the Committee noted that the USPSTF recommendation has been changed to a "B" level but agreed that the underlying evidence presented appears to be directionally the same since the last NQF endorsement review.
- The Committee highlighted that only 38% of the visits in one cohort in 2014 had appropriate testing, signaling a significant gap in care.
- The Committee expressed concerns about the exclusive focus on women and the unintended consequences for not including men in the measure. The developer clarified that the Task Force evaluated this before this measure was originally approved and the evidence for a direct health benefit was limited to women. The Committee highlighted that the 2014 USPSTF recommendation acknowledged the importance of men in this population, citing extensively the CDC recommendations in screening and treating men but recognized the limitation of data.

Changes to evidence from last review

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

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

Updates:

- The developer noted that the 2014 USPSTF found little direct evidence on the effectiveness of screening for chlamydia in men or low-risk women. Although no additional evidence was provided for the measure, beyond the updated Centers for Disease Control and Prevention (CDC). 2021. "Sexually Transmitted Diseases: Chlamydia—CDC Fact Sheet." http://www.cdc.gov/std/chlamydia/STDFact-chlamydia-detailed.htm, developers elaborated on questions posed in the 2014 USPSTF evidence, yet do not discuss current available evidence for these four questions:
  - How effective is screening for gonorrhea and chlamydia in reducing complications of infection and transmission or acquisition of disease in asymptomatic, sexually active men and nonpregnant women, including adolescents?
  - How effective are different screening strategies in identifying persons with gonorrhea and chlamydia?
  - How accurate are screening tests in detecting gonorrhea and chlamydia?
  - What are the harms of screening for gonorrhea and chlamydia?
- The evidence in the 2015 CDC guidelines state anyone under 24 years who are sexually active should be screened for chlamydia, two-thirds of new chlamydial infections occur among youth aged 15-24 years, and that it is estimated that 1 in 20 sexually active young women aged 14-24 years has chlamydia.

Questions for the Committee:

- Is the evidence directly applicable to the process of care being measured?
- The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?

Guidance from the Evidence Algorithm

- Process measure (Box 3) → Systematic review with QQC (Box 4) → Moderate certainty that the net benefit is substantial (Box 5b) → Moderate

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provided performance data from HEDIS database on the rate of chlamydia screening for women. The data was stratified by age, year and health plan type commercial or Medicaid.
  - **Commercial**
    - 16 to 20 years of age (2017, 2018, 2019): mean (43-44%); SD (12-13%); performance at 90th percentile (60-63%)
    - 21 to 24 years of age: (2017, 2018, 2019): mean (53-55%); SD (10%); performance at 90th percentile (67-70%)
    - Total: (2017, 2018, 2019): mean (48-50%); SD (11%); performance at 90th percentile (64-66%)
  - **Medicaid**

    - 16 to 20 years of age (2017, 2018, 2019): mean (54-55%); SD (12-13%); performance at 90th percentile (70-71%)
    - 21 to 24 years of age: (2017, 2018, 2019): mean (63-64%); SD (10%); performance at 90th percentile (74-75%)
    - Total: (2017, 2018, 2019): mean (58%); SD (11-12%); performance at 90th percentile (71-72%)
- The variations and low performance show a gap in care and an opportunity for improvement.

Disparities

- The developer noted the literature shows racial/ethnic differences in screening rates (six times higher in Black women compared to White women) and infection rates (higher in Black, American Indian/Alaska Native (AI/AN) and Native Hawaiian/Other Pacific Islander (NHOPI) populations).
- Developer does not currently collect or stratify performance data by race, ethnicity, or language.
- The developer noted that HEDIS data are stratified by type of insurance (e.g., Commercial, Medicaid, Medicare), but could be stratified by demographic variables. They also note that the HEDIS measures *Race/Ethnicity Diversity of Membership* and *Language Diversity of Membership* can be used to assess disparities in the health plan population. A complement electronically specified clinical quality measure (eCQM) (i.e., CMS153v), not endorsed by NQF, is in current use suggesting the availability of supplemental disparities data elements, including race, ethnicity, payer, and sex for stratification.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- If no disparities information is provided, are you aware of other evidence that disparities exist in this area of healthcare?
- Based on age differences in the varied guidelines and identified disparities based on race and ethnicity, should the lower age parameter appropriate to existing evidence?
- Are other populations appropriate for denominator inclusion, specifically all pregnant women in the first trimester, men, or men have sex with men (MSM)?

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

Committee Pre-evaluation Comments:

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

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- NO
- Not aware of any new studies/information that changes the evidence base for this measure.
- Strong systematic review and evidence
- Moderate
- Measures conducted by system review and empirical data review. The evidence supports the measures
- I wondered if the age range for this measure (16-24 years) had been tested recently and found to still be relevant. A study from Nov. 2020 found a decreasing incidence of chlamydia in the youngest age range (in this study, age 12-17) but an increase in positivity in the 18-24 year old and 25-30 year old groups. (https://www.sciencedirect.com/science/article/pii/S0749379720302713). I would like the measure developer to comment on the risk/benefit of increasing the recommended screening age.
- Yes, evidence supports this process measure. It applies directly and the process allows for early detection and treatment.
- Significant disease burden

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

- yes
- Given many organizations' explicit focus on health equity, including disparities in care that stem from implicit provider bias and structural racism, as well as documented racial/ethnic differences in screening rates, it feels critical for the developer to collect and stratify performance data by race/ethnicity. In addition, based on current guidelines, a lower age parameter - 15 years (in alignment with USPSTF recommendations) or ~11 years (as CDC guidelines recommend that all sexually active women younger than 25 years be tested for chlamydia every year) – is more appropriate for this measure. The 16-year-old lower age parameter feels arbitrary and may lead to missed opportunities for screening younger populations who are at risk for chlamydia.
- moderate gap with demonstrated disparities
- There is a substantial performance gaps among different populations that needs to be addressed
- The developer provided performance data from HEDIS database on the rate of chlamydia screening for women. The data was stratified by age, year and health plan type commercial or Medicaid
- The data show a gap in care that warrants a national performance measure. Data should be stratified by race, ethnicity, and payer. Discuss inclusion of trans-men and trans-women.
- Yes, current data was provided and gap in care identified to warrant a national performance measure. Disparities were noted from literature reviews.
- yes gap yes disparities

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

2c. For composite measures: empirical analysis support composite approach

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

Complex measure evaluated by Scientific Methods Panel? ☐ **Yes** ☒ **No**

Evaluators: NQF Staff

Reliability

2a1. Specifications

- Performance assesses outpatient: clinic and office visit types, pharmacy, and laboratory claims to determine the denominator population defined by age and identifying for sexual activity. The developers define pharmacy data (see Contraceptives Medications List) and claim/encounter data (Pregnancy Value Set, Sexual Activity Value Set, and Pregnancy Tests Value Set).
- The identification of sexual activity and chlamydia testing does not require testing to be performed after identification, just during the same measurement period. This could lower performance in providers in identifying sexual activity late in the measurement period.
- The use of over the counter (OTC) pregnancy tests and pregnancy prophylaxis is not apparent in the denominator codes.
- Confidential OTC chlamydia testing and treatment products are not apparent in the data sets.
- The use of sexual activity, testing and treatment coding from inpatient, emergency, and other care settings is not clearly defined in the data sets.
- The 2016 level of analysis (LOA) was Health Plan and Integrated Delivery System and the current LOA is Health Plan only. The MIPS use of this measure is used in Integrated Care Delivery Value-Based Payment (VBP) Alternate Payment Models (APMs).
- The 2016 CDC guidelines recommend a pregnancy test within 1 week of surgery. This is not identified as a denominator exclusion.

2a2. Reliability Testing

- 2019 HEDIS Health Plan performance data was used to assess accountable entity/measure score reliability in 402 commercial plans and 251 Medicaid plans in reported diverse geographies and sizes. This Medicaid and commercial plan types serves as a proxy for income and other socioeconomic factors. For this testing, only age and plan type were provided.
- The developer used a beta-binominal model to assess the signal-to-noise ratio. Using this method, the total mean commercial reliability score was 0.979 and the mean Medicaid reliability score was 0.984.
- The developer also reported the following scores stratified by age.
    - Commercial
        - Chlamydia Screening in Women (16-20): 0.975
        - Chlamydia Screening in Women (21-24): 0.964
        - Chlamydia Screening in Women (Total): 0.979
    - Medicaid
        - Chlamydia Screening in Women (16-20): 0.978
        - Chlamydia Screening in Women (21-24): 0.955
        - Chlamydia Screening in Women (Total): 0.984
- For signal to noise, a minimum reliability score of 0.7 is used to show sufficient signal strength to discriminate performance between accountable entities. The testing suggests that all indicators within this measure have good reliability between 0.7 and 1.0. The signal-to-noise total findings are higher in both Commercial and Medicaid populations than the age stratifications.
- The developer also provided Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) for the Chlamydia Screening in Women (Total Ages 16-24) Measure by Terciles of the Denominator Size and for All Submissions Stratified by Plan Type in Calendar Year 2019 Data, and Distribution of Plan-Level Signal-To-Noise Reliability for the Chlamydia Screening in Women (Total Ages 16-24) measure by Tertiles of the Denominator Size and for All Submissions by Plan Type in Calendar Year 2019 Data.

Validity

- The developer performed validity testing at the accountable entity/measure score level through construct validity testing and face validity.
- The developer conducted Pearson correlation for construct validity against NCQA's *Cervical Cancer Screening* measure.
- Results:
    - Hypothesis: Positive Correlation with *Cervical Cancer Screening*
        - Commercial
            - Women (16-20): Correlation coefficient = 0.53, $p < 0.001$
            - Women (21-24): Correlation coefficient = 0.53, $p < 0.001$
            - Combines age totals were not provided.
        - Medicaid
            - Women (16-20): Correlation coefficient = 0. 32, $p < 0.001$
            - Women (21-24): Correlation coefficient = 0.44, $p < 0.001$
            - Combines age totals were not provided.
    - The developer concluded that there is a moderate correlation between this measure and the *Cervical Cancer Screening* measure.
    - The developer noted the correlation is weaker for the younger population because the comparison measure does not include that population and concluded that correlation could not be considered high due to that factor.

- The developer currently allows health plans to apply exclusions to their results and do not collect data on exclusions for HEDIS measure reporting. They report assessing and validating exclusions applied to the eligible denominator with reporting quantitative or qualitative findings. Current exclusions are expert consensus recommendations only. Empirical testing was not provided.
- Meaningful differences in performance were calculated using inter-quartile ranges (IQR) for each indicator as a measure of performance dispersion and interpreted as the difference between the 25th and 75th percentile. Commercial plan IQRs were 16%, 13%, and 14% ranging from 35-61%, and Medicaid IQRs were 16%, 10%, and 15% ranging from 47%-70% for ages 16-20, 21-24, and total years, respectively. All p-values were <0.001.
- The developer noted that face validity was also conducted and referred to the 2016 face validity results in the 2021 submission. The developer convened a panel of 33 members to assess the face validity of this measure. Results of the face validity testing were not provided in 2016 or in this submission. The developer reported that panel and public found the measure to be valid.
- A sound description of assessing for "material bias" in missing data is provided without quantitative or qualitative findings.
- The developer provided a sound description for March 2011 and March 2012 ICD-10 coding conversion. No additional details were provided on subsequent ICD-10 updates and approval processes.

Questions for the Committee regarding reliability:

- Does the Committee have specification concerns related to testing timing, coding questions, or OTC approach of sexual activity, testing, and treatment identification and reliably assessing performance?
- What prompted the LOA change from Health Plan and Integrated Delivery System in 2016 to Health Plan only in the current submission?
- Does the Committee have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The signal-to-noise total findings are higher in both Commercial and Medicaid populations than the age stratifications. The staff is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Does the Committee have any concerns with the Pearson results in lower volume Medicaid plans, the developer's assessment of moderate if 0.25 to 0.75, or the lack of total findings for construct validity to the Cervical Cancer Screening measure (Tables 4 and 5 of the testing attachment)?
- Does the Committee have any concerns regarding the validity of the measure (e.g., lack of quantitative and qualitative findings for exclusions analysis, missing data, or ICD-10/updates testing)? Do measures with claims-only data prevent the developer from providing comprehensive threats to validity analyses?
- Overall, the staff is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

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

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- none
- How is "sexual activity" defined and what is the developer doing to ensure this definition is applied consistently when collecting performance data? Differences in screening for sexual activity (e.g., what behaviors constitute "sexual activity"? how far back does the patient's last sexual encounter have to be for them to be considered as “currently sexually active”) could result in inconsistent measurement and threats to reliability.
- no concerns
- High reliability
- The developer used a beta-binominal model to assess the signal-to-noise ratio. Using this method, the total mean commercial reliability score was 0.979 and the mean Medicaid reliability score was 0.984. I do not have concerns.
- As the use of OTC testing and emergency contraception increases, how will that affect the reliability of this measure, which is based on claims data?
- Data elements are clearly defined. All steps are clear. No concerns.
- I am concerned that data collection is not complete and is flawed in that the identification of sexual active <24 years is not complete and may capture some non-sexual active by mistake. I am also concerned re the focus on women only

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- none
- No concerns.
- no concerns
- No
- No
- None
- No
- test for STD in minors may be in types of encounter that are not easily picked up by administrative data.

2b1. Validity - Testing: Do you have any concerns with the testing results?

- none
- No concerns.
- no concerns
- No
- No
- None
- No
- see above

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- none

- No concerns.
- no concerns
- No real threat from missing data
- No threats to validity
- No concerns.
- Unknown, HEIDS does address missing data through an audit process.
- missing data is a threat. No way to assess how much data is missing.

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment) 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- no concerns
- No comment.
- no concerns-
- Exclusions are appropriate
- No concerns
- I would like to understand how the developer concludes that allowing plans to apply exclusions to their results through expert consensus recommendations is not a concern and a threat to validity.
- Social risk factor data was not available in the reporting. There are actions underway to incorporate social risk factors.
- nothing to add

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

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

- This measure is based on administrative claims data that is generally considered to be very feasible and low burden.
- Data elements for this measure are coded by someone other than person obtaining original information (e.g., CPT and ICD-10 codes) on claims.
- Per developer, data is collected during the provision of care and coded by someone other than person obtaining original information and all data elements are in defined fields in a combination of electronic sources.
- NCQA conducts audits for all HEDIS collection and reporting processes. NCQA conducts an independent audit of HEDIS process to verify integrity of HEDIS collection and reporting system. NCQA also uses Policy Clarification Support System to generate ongoing feedback from measure users.
    - Per developer, NCQA goals align with NQF that noncommercial uses do not require the consent of the measure developer. An example of such would be use by health care physicians in connection with their own practice. However, commercial use of the measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product

or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?

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

Committee Pre-evaluation Comments:

Criteria 3: Feasibility

3. **Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?**

- feasible measure
- No concerns.
- highly feasible
- No concerns. All of the data can be part of electronic data capture
- No concerns
- No concerns.
- no concerns
- feasible except that confidential encounters for minors would not likely be picked up.

Criterion 4: Usability and Use

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

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

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

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

- This measure is used in the following public reporting and accountability programs: California Align. Measure. Perform (AMP) Commercial HMO Program, California AMP Medi-Cal Managed Care Program, Medicaid Adult Core Set, NCQA Health Plan Rating/Report Cards, NCQA State of Health Care Annual Report, NCQA Health Plan Accreditation, NCQA Accountable Care Organization Accreditation, NCQA Quality Compass, and the Qualified Health Plan Quality Rating System.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- The developer noted that feedback on this measure has focused on the definitions of "sexually active" and clarifications around whether direct optical observation would count as screening.

Additional Feedback:

- N/A

***Questions for the Committee**:*

- How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

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

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

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

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- The developer reported that commercial health plan measure scores have improved approximately 1% over the last three years and the performance of Medicaid plans has remained consistent.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- The developer stated, there were no identified unintended consequences for this measure during testing or since implementation. No quantitative or qualitative findings are provided.

Potential harms

- The developer states there were no identified potential harms for this measure during testing or since implementation. No quantitative or qualitative findings are provided.

Additional Feedback

- Feedback from previous endorsement cycle indicate that MAP committee recommended the developer explore creating a composite of all COPD measures and then link that composite with the COPD resource use measure.

***Questions for the Committee**:*

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

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

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided? 4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- no concerns
- No concerns.
- accountable- yes!
- Yes
- No concerns
- No concerns.
- Publicly available data.
- concerned that over testing will occur

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- no concerns
- No concerns - believe any potential harm cause by false-positive tests is outweighed by benefits of increased rates of routine screening.
- highly usable
- None
- No concerns
- No concerns.
- The harms are very low rates or false-positive or false-negative results. The benefits outweigh any harm.
- over testing could lead to family consequences for minors who might have to explain why they were tested to parents

Criterion 5: Related and Competing Measures

Related or competing measures

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

Harmonization

- The developer reports that these measures are harmonized to extent possible.

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- no
- No.
- #409- harmonized to extent possible-
- Yes. But these measures have bene harmonized
- no concerns
- No concerns.
- The competing measure is for those 13 years + with HIV/AIDs so it's a different denominator.
- none

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 06/29/2021

- No NQF Members have submitted support/non-support choices as of this date.
- No Public or NQF Member comments submitted as of this date.

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0033

Measure Title: Chlamydia Screening in Women (CHL)

Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**

☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**

☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☐ **Facility** ☒ **Health Plan**

☐ **Population: Community, County or City** ☐ **Population: Regional and State**

☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

Submission document: "MIF_xxxx" document, items S.1-S.22

***NOTE**: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**
    - Performance assesses outpatient: clinic and office visit types, pharmacy, and laboratory claims to determine the denominator population defined by age and identifying for sexual activity. The developers define pharmacy data (see Contraceptives Medications List) and claim/encounter data (Pregnancy Value Set, Sexual Activity Value Set, and Pregnancy Tests Value Set). The numerator clinical action assesses submitted laboratory claims that identify completion of a chlamydia testing. Is data from
    - The identification of sexual activity and chlamydia testing does not require testing to be performed after identification, just during the same measurement period. This could lower performance in providers in identifying sexual activity late in the measurement period.
    - The use of over the counter (OTC) pregnancy tests and pregnancy prophylaxis is not apparent in the denominator codes.
    - Confidential OTC chlamydia testing and treatment products are not apparent in the data sets.
    - The use of sexual activity, testing and treatment coding from inpatient, emergency, and other care settings is not clearly defined in the data sets.
    - The 2016 level of analysis (LOA) was Health Plan and Integrated Delivery System and the current LOA is Health Plan only. The MIPS use of this measure is used in Integrated Care Delivery Value-Based Payment (VBP) Alternate Payment Models (APMs).
    - The 2016 CDC guidelines recommend a pregnancy test within 1 week of surgery. This is not identified as a denominator exclusion.

RELIABILITY: TESTING

Submission document: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?
   ☐ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2
    - The developer conducted updated performance measure score reliability testing using 2019 claims data at the health plan level with Commercial and Medicaid plans. Reliability testing was performed by using a beta-binomial model (i.e., signal to noise), Standard Error (SE), and 95% Confidence Intervals (95% CI).
7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3
    - The developer reported a mean commercial reliability score of 0.979 and a mean Medicaid reliability score of 0.984. The developer reported a mean commercial score for women aged 16-20 of 0.975 and for ages 21-24 a score of 0.964. The developer reported a mean commercial score for women aged 16-20 of 0.978 and for ages 21-24 a score of 0.955. The signal-to-noise total findings are higher in both Commercial and Medicaid populations than the collective age stratifications.

- According to Adams, et al., for signal to noise, a reliability of 0.70 – 0.80 is generally considered the acceptable threshold for reliability, 0.80 – 0.90 is considered high reliability, and 0.90 – 1.0 is considered very high.
- The developer also provided Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) for the Chlamydia Screening in Women (Total Ages 16-24) Measure for 2019 Plan Data, and Distribution of Plan-Level Signal-To-Noise Reliability for the Chlamydia Screening in Women (Total Ages 16-24) measure by Tertiles for All Submissions for 2019 Plan Data.
- The developer concludes that both the reliability results for both populations had high reliability.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

   ☒ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - Precise specifications (Box 1) → empirical reliability testing (box 2) →used computed performance scores for measure entities (Box 4) → Appropriate method used (Box 5) →High reliability statistic and scope (Box 6a) → High

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

    Submission document: Testing attachment, section 2b2.

    - The following populations are excluded from this measure:
      Women who received a pregnancy test to determine contraindications for medication (isotretinoin) or x-ray and women who were in hospice or using hospice services during the measurement year.
    - Were pre-surgical pregnancy tests considered as a measure exclusion as stated in the 2016 CDC guidelines?
    - Empirical validity testing was not conducted for exclusions and missing data/material biases. Empirical findings were not provided for the 2011-2012 ICD-10 conversion or face validity.

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

Submission document: Testing attachment, section 2b4.

- The developer calculated an inter-quartile range (IQR) for each type of plan.
- For commercial plans, the IQR for the 16-24 age range was 14%, which represents an average of 1,236 additional women receiving chlamydia screening in the 25th percentile compared to the 75th percentiles.
- For Medicaid plans, the IQR for the 16-24 age range was 15%, which represents an average of 988 additional women receiving chlamydia screening in the 25th percentile compared to the 75th percentiles.
- No concerns.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Testing attachment, section 2b5.
    - This measure does not use multiple data sources.
    - No concerns.
15. **Please describe any concerns you have regarding missing data.**
Submission document: Testing attachment, section 2b6.
    - The developer reported that missing data in HEDIS measures are addressed in a structured way through an audit process to ensure the eligible population and numerator events for each measure are correctly identified and reported. No empirical validity testing is provided.
16. **Risk Adjustment**

16a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☐ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

16d. **Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

16d.5. Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

16e**. Assess the risk-adjustment approach**

VALIDITY: TESTING

17. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**
18. **Method of establishing validity of the measure score:**

☒ **Face validity**

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

☐ **N/A (score-level testing not conducted)**

19. **Assess the method(s) for establishing validity**

Submission document: Testing attachment, section 2b2.2

- The developer provided updated 2019 testing information. Empirical validity testing at the score level was conducted through construct validity to determine whether there was a correlation between: Chlamydia Screening and Cervical Cancer screening.
- The developer hypothesized that organizations that performed well on this measure would perform well on this measure (i.e., positive correlation).
- The developer stated that the correlation was considered high (strong) if the correlation coefficient is 0.75 to 1, moderate if 0.25 to 0.75, and low (weak) if 0 to 0.25.
- The submission also described face validity testing for new measures. No empirical findings were provided.

20. **Assess the results(s) for establishing validity**

Submission document: Testing attachment, section 2b2.3

- The developer reported the following results:
    - Pearson correlation coefficients were 0.53 (16-20 and 21-24) for Commercial plans and 0.32 (16-20) and 0.44 (21-24) in Medicaid plans. Combines age totals were not provided.
    - The developer concluded that the results suggested that the correlations were moderate, i.e., suggest that plans that perform well on this measure are moderately likely to perform well on the other measures.

21. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

☐ **Not applicable** (score-level testing was not performed)

22. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**
NOTE that data element validation from the literature is acceptable.

Submission document: *Testing attachment, section 2b1.*

☐ **Yes**

☐ **No**

☒ **Not applicable** (data element testing was not performed)

23. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☐ **Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

24. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

- Potential threats to validity relevant to the measure were empirically assessed (Box 1) → Empirical validity testing was conducted (Box 2) → Validity testing was conducted (Box 5) → Correlation of performance measure scores conducted and reported (Box 6) → Moderate, correlation reported (7b) → MODERATE

ADDITIONAL RECOMMENDATIONS

25. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**
    - Empirical testing was not conducted for all threats to validity.

26. **ding Committee? If so, please list those concerns below.**

Developer Submission

NQF #: 0033

Corresponding Measures:

De.2. Measure Title: Chlamydia Screening in Women (CHL)

Co.1.1. Measure Steward: National Committee for Quality Assurance

De.3. Brief Description of Measure: The percentage of women 16–24 years of age who were identified as sexually active and who had at least one test for chlamydia during the measurement year.

1b.1. Developer Rationale: This measure assesses the percentage of women 16-24 years of age who were identified as sexually active and who received a test for chlamydia. The improvement in quality envisioned by the use of this measure is increased identification of untreated chlamydia infections in women that can lead to serious and irreversible complications and can be unknowingly transmitted to sexual partners. Despite the availability of effective treatments, a large proportion of sexually active individuals continue to go undiagnosed due to the disease's asymptomatic nature. Early detection, screening, and treatment have proven to be effective in managing and preventing chlamydia.

Centers for Disease Control and Prevention (CDC). 2021. "Sexually Transmitted Diseases: Chlamydia—CDC Fact Sheet." http://www.cdc.gov/std/chlamydia/STDFact-chlamydia-detailed.htm

S.4. Numerator Statement: Women who were tested for chlamydia during the measurement year.

S.6. Denominator Statement: Women 16-24 years of age who had a claim or encounter indicating sexual activity.

S.8. Denominator Exclusions: Women who received a pregnancy test to determine contraindications for medication (isotretinoin) or x-ray.

Women who were in hospice or using hospice services during the measurement year.

De.1. Measure Type: Process

S.17. Data Source: Claims, Enrollment Data

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Oct 25, 2016

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A

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

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

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

nqf_evidence_attachment_7.1_508Compliant.docx

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

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

No

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0033

Measure Title: Chlamydia Screening in Women

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 4/9/2021

1a.1. This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Outcome:

☐Patient-reported outcome (PRO):

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*):

☒ Process: **Chlamydia Screening**

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

2021 Submission:

Sexually active women >>> increased risk of chlamydial infection>>> screening for chlamydia occurs>>> positive chlamydia test result>>> treatment>>> decreased incidence of pelvic inflammatory disease, infertility, and perinatal infections.

2016 Submission:

There is good evidence that screening for Chlamydial infection in women who are at increased risk can reduce the incidence of pelvic inflammatory disease (PID), infertility and perinatal infections. The US Preventive Services Task Force (USPSTF) concluded that the benefits of screening women at increased risk are substantial.

1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☒ US Preventive Services Task Force Recommendation

☒ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	2021 Submission No changes 2016 Submission LeFevre, M.L. Screening for Chlamydia and Gonorrhea: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2014;161(12):902-10. Guideline available from: http://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/chlamydia-and-gonorrhea-screening, accessed February 16, 2016.
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	2021 Submission: No changes 2016 Submission "The USPSTF (2014) recommends screening for chlamydia in sexually active females aged 24 years or

Systematic Review	Evidence
	younger and in older women who are at increased risk for infection."
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	2021 Submission: No changes 2016 Submission The USPSTF concludes with moderate certainty that screening for chlamydia is associated with moderate net benefit in all sexually active women aged 24 years or younger and in older women who are at increased risk for infection.
Provide all other grades and definitions from the evidence grading system	N/A
Grade assigned to the **recommendation** with definition of the grade	2021 Submission: No changes 2016 Submission Grade: B Recommendation. Grade B: The USPSTF recommends the services. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.
Provide all other grades and definitions from the recommendation grading system	2021 Submission No changes 2016 Submission Grade A: The USPSTF recommends the services. There is high certainty that the net benefit is substantial. Grade C: The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is moderate or high certainty that the net benefit is small. Grade D: The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harm outweighs the benefits. I Statement: The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking,

Systematic Review	Evidence
	of poor quality, or conflicting and the balance of benefits and harms cannot be determined.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	2021 Submission In 2014, the US Preventive Services Task Force and the Agency for Healthcare Research and Quality determined four Key Questions for men and nonpregnant women and identified, reviewed and rated the quality of studies published since the previous 2007 evidence review. Key Questions were also identified for pregnant women, but no studies addressing screening in pregnant women met inclusion criteria. *Men and non-pregnant women:* Key Question 1: How effective is screening for gonorrhea and chlamydia in reducing complications of infection and transmission or acquisition of disease in asymptomatic, sexually active men and nonpregnant women, including adolescents? • 1 RCT of Immediate testing and treatment vs. deferred screening. The USPSTF review found that "the POPI trial was a good-quality RCT of 2,529 sexually active young women (mean age, 21 years [range, 16 to 27 years]) recruited from universities and colleges in the United Kingdom." Key Question 2: How effective are different screening strategies in identifying persons with gonorrhea and chlamydia? • No studies Key Question 3: How accurate are screening tests in detecting gonorrhea and chlamydia? • 8 Clinical Trials for comparisons of multiple screening tests against reference standards. "Ten new fair-quality diagnostic accuracy studies reporting test characteristics of FDA-cleared NAATs met inclusion criteria, including six for gonorrhea and eight for chlamydia." Key Question 4: What are the harms of screening for gonorrhea and chlamydia? • 6 Clinical Trials for comparisons of multiple screening tests against reference standards. The USPSTF found that four of the studies

Systematic Review	Evidence
	were of fair quality while the remaining two were good quality. 2016 Submission Non-pregnant women at increased risk. There is good evidence that screening for Chlamydial infection in women who are at increased risk can reduce the incidence of pelvic inflammatory disease (PID). The US Preventive Services Task Force (USPSTF) concluded that the benefits of screening women at increased risk are substantial. Pregnant women at increased risk. There are no studies evaluating the effectiveness of screening for chlamydial infection in pregnant women who are at increased risk. The USPSTF, however, found the following: 1) screening identifies infection in asymptomatic pregnant women; 2) there is a relatively high prevalence of infection among pregnant women who are at increased risk; and 3) there is fair evidence of improved pregnancy and birth outcomes for women who are treated for chlamydial infection. The USPSTF concluded that the benefits of screening pregnant women who are at increased risk are substantial. Women not at increased risk. The USPSTF identified no studies documenting the benefits of screening women, including pregnant women, who are not at increased risk for chlamydial infection. While recognizing the potential benefit to women identified through screening, the USPSTF concluded the overall benefit of screening would be small, given the low prevalence of infection among women not at increased risk. Men. While concluding that the direct benefit to men of screening was likely to be small, the USPSTF noted that screening for chlamydial infection in men may be beneficial if it were to lead to a decreased incidence of chlamydial infection in women. The USPSTF did not, however, find evidence to support this outcome, and therefore concluded that the benefits of screening men are unknown. The USPSTF identified this as a critical gap in the evidence.
Estimates of benefit and consistency across studies	2021 Submission The USPSTF found adequate direct evidence that screening reduces complications of chlamydial infection in women who are at increased risk, with a

Systematic Review	Evidence
	moderate magnitude of benefit. Previous USPSTF reviews identified 2 RCTs of the effectiveness of screening for chlamydia for the prevention of PID in nonpregnant women at increased risk for infection; in 1 large RCT, a strategy of identifying, testing, and treating women at increased risk for cervical chlamydial infection was associated with significantly reduced incidence of PID (relative risk [RR], 0.44 [95% CI, 0.20 to 0.90]). The 2014 USPSTF review identified 1 good-quality RCT and among asymptomatic women, 0.6% in the screening group versus 1.6% in the deferred group developed PID during follow-up (RR, 0.39 [CI, 0.14 to 1.08]). Study limitations may have attenuated intervention effects and the study may have been underpowered. The USPSTF found little direct evidence on the effectiveness of screening for chlamydia in men or low-risk women. It found that the overall prevalence of chlamydial infection in the general population varies widely depending on age and other risk factors. 2016 Submission Consistent The USPSTF determined there was a positive net benefit.

What harms were identified?	**2021 Submission** **Potential harms of screening for chlamydia include false-positive or false-negative results as well as labeling and anxiety associated with positive results. The USPSTF found adequate evidence that the harms of screening for chlamydia are small to none.** **2016 Submission** **The USPSTF determined there was a positive net benefit.**
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	2021 Submission The USPSTF is conducting an updated evidence review and is expected to release updated clinical guideline recommendations for chlamydia screening in 2021.

1a.4 OTHER SOURCE OF EVIDENCE

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

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1b. Performance Gap

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

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

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

If a COMPOSITE *(e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.*

This measure assesses the percentage of women 16-24 years of age who were identified as sexually active and who received a test for chlamydia. The improvement in quality envisioned by the use of this measure is increased identification of untreated chlamydia infections in women that can lead to serious and irreversible complications and can be unknowingly transmitted to sexual partners. Despite the availability of effective treatments, a large proportion of sexually active individuals continue to go undiagnosed due to the disease's asymptomatic nature. Early detection, screening, and treatment have proven to be effective in managing and preventing chlamydia.

Centers for Disease Control and Prevention (CDC). 2021. "Sexually Transmitted Diseases: Chlamydia—CDC Fact Sheet." http://www.cdc.gov/std/chlamydia/STDFact-chlamydia-detailed.htm

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis*. (**This is required for maintenance of endorsement**. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

The following data are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. Performance data is summarized at the health plan level and summarized by mean, standard deviation, minimum health plan performance, maximum health plan performance and performance at 10th, 25th, 50th, 75th, and 90th percentile. Data is stratified by year and product line (i.e., commercial, Medicaid).

The following data demonstrate the variation in the rate of chlamydia screening for women across health plans. These gaps in performance underscore the opportunity for improvement.

Chlamydia Screening in Women

Commercial Rate – 16 to 20 years of age

YEAR| N | MEAN | ST DEV | Min | 10TH | 25TH | 50TH | 75TH | 90TH | Max

2019 | 393 | 44% | 13% | 21% | 30% | 35% | 42% | 51% | 63% | 82%

2018 | 382 | 43% | 12% | 21% | 30% | 35% | 41% | 50% | 62% | 81%

2017 | 386 | 43% | 12% | 18% | 30% | 35% | 40% | 79% | 60% | 75%

Commercial Rate – 21 to 24 years of age

YEAR| N | MEAN | ST DEV | Min | 10TH | 25TH | 50TH | 75TH | 90TH | Max

2019 | 396 | 55% | 10% | 26% | 43% | 48% | 54% | 61% | 70% | 83%

2018 | 384 | 54% | 10% | 29% | 42% | 47% | 53% | 59% | 69% | 80%

2017 | 385 | 53% | 10% | 0% | 41% | 46% | 51% | 59% | 67% | 80%

Commercial Rate - Total

YEAR| N | MEAN | ST DEV | Min | 10TH | 25TH | 50TH | 75TH | 90TH | Max

2019 | 402 | 50% | 11% | 23% | 36% | 42% | 48% | 56% | 66% | 82%

2018 | 391 | 49% | 11% | 25% | 37% | 42% | 47% | 55% | 65% | 80%

2017 | 390 | 48% | 11% | 0% | 36% | 41% | 46% | 54% | 64% | 77%

Medicaid Rate – 16 to 20 years of age

YEAR| N | MEAN | ST DEV | Min | 10TH | 25TH | 50TH | 75TH | 90TH | Max

2019 | 247 | 55% | 13% | 12% | 38% | 47% | 54% | 63% | 70% | 88%

2018 | 213 | 55% | 12% | 28% | 40% | 47% | 54% | 63% | 71% | 88%

2017 | 217 | 54% | 12% | 12% | 40% | 47% | 53% | 63% | 70% | 91%

Medicaid Rate – 21 to 24 years of age

YEAR| N | MEAN | ST DEV | Min | 10TH | 25TH | 50TH | 75TH | 90TH | Max

2019 | 236 | 64% | 10% | 16% | 55% | 60% | 65% | 70% | 74% | 87%

2018 | 216 | 64% | 10% |3 1% | 50% | 58% | 65% | 70% | 75% | 84%

2017 | 222 | 63% | 10% | 14% | 51% | 57% | 64% | 70% | 74% | 82%

Medicaid Rate - Total

YEAR| N | MEAN | ST DEV | Min | 10TH | 25TH | 50TH | 75TH | 90TH | Max

2019 | 251 | 58% | 12% | 14% | 43% | 51% | 58% | 66% | 71% | 86%

2018 | 218 | 58% | 11% | 31% | 44% | 50% | 58% | 66% | 72% | 84%

2017 | 224 | 58% | 11% | 13% | 45% | 51% | 56% | 65% | 71% | 87%

The data references are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. It includes the number of health plans included in HEDIS data collection and the median denominator for the measure across health plans.

Commercial – 16 to 20 years of age

YEAR | N Plans | Median Denominator Size per plan

2019 | 393| 1,385

2018 | 382| 1,469

2017 | 386| 1,374

Commercial – 21 to 24 years of age

YEAR | N Plans | Median Denominator Size per plan

2019 | 396| 1,665

2018 | 384| 1,713

2017 | 385| 1,652

Commercial – Total

YEAR | N Plans | Median Denominator Size per plan

2019 | 395| 3,011

2018 | 391| 3,062

2017 | 390| 2,925

Medicaid – 16 to 20 years of age

YEAR | N Plans | Median Denominator Size per plan

2019 | 247| 2,261

2018 | 213| 2,697

2017 | 217| 2,667

Medicaid – 21 to 24 years of age

YEAR | N Plans | Median Denominator Size per plan

2019 | 236| 1,482

2018 | 216| 1,663

2017 | 222| 1,859

Medicaid – Total

YEAR | N Plans | Median Denominator Size per plan

2019 | 251| 3,796

2018 | 218| 4,531

2017 | 224| 4,582

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

N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. ***(This is required for maintenance of endorsement****. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

HEDIS data are stratified by type of insurance (e.g., Commercial, Medicaid). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities if the data are available to a plan. NCQA is actively engaged with partners including the CMS Office of Minority Health in identifying feasible methods to further integrate social risk factors into health plan quality measures, with a focus on stratification. Our work is aligned with recent recommendations from MedPAC and ASPE on optimal methods for addressing social risk in quality measurement and programs. 1,2 This is an NCQA wide initiative. Our intent is to implement methods to bridge data concerns in the future.

HEDIS includes two measures that can be used as tools for assessing race/ethnicity and language needs of a plan's population: Race/Ethnicity Diversity of Membership and the Language Diversity of Membership. These measures promote standardized methods for collecting these data and follow Office of Management and Budget and National Academy of Medicine guidance for collecting and categorizing race/ethnicity and

language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing, and using race/ethnicity and language data to assess health care disparities.

1. Medicare Payment Advisory Commission. (2020). The Medicare Advantage program: Status report. In Report to the Congress: Medicare Payment Policy (p. 397). http://medpac.gov/docs/default-source/reports/mar20_medpac_ch13_sec.pdf
2. Office of the Assistant Secretary for Planning and Evaluation, & U.S. Department of Health & Human Services. (2020). Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-basedpurchasing-programs

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

Studies show that racial/ethnic disparities continue to exist in chlamydia infection rates, particularly in Black, American Indian/Alaska Native (AI/AN) and Native Hawaiian/Other Pacific Islander (NHOPI) populations. In 2018, the rate of reported chlamydia infections for Black females was 5 times that of white females (1,411.1 and 281.7 cases per 100,000 population, respectively). The rate among AI/ANs and NHOPIs were 3.7 times and 3.3 times the rate among Whites, respectively (784.8 cases per 100,000 population and 700.8 cases per 100,000 population) (Centers for Disease Control and Prevention, 2019). Rates of screening also differ by race/ethnicity showing disparities. One study found that chlamydia screening rates for women aged 15-25 were 45.6% for white women and 57.5% for black women (Patel, 2016).

Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2018. Atlanta: U.S. Department of Health and Human Services; 2019. DOI: 10.15620/cdc.79370.

Patel CG, Chesson HW, Tao G. Racial Differences in Receipt of Chlamydia Testing Among Medicaid-Insured Women in 2013. Sex Transm Dis. 2016;43(3):147-151. doi:10.1097/OLQ.0000000000000405

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

De.5. Subject/Topic Area *(check all the areas that apply):*

Infectious Diseases (ID) : Sexually Transmitted

De.6. Non-Condition Specific*(check all the areas that apply):*

Screening

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

Children, Women

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

NA

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

This is not an eMeasure **Attachment:**

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

Attachment **:** 033_CHL_Spring_2021_Value_Sets-637553860316459511.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

As part of NCQA's annual measure maintenance, we routinely make coding and other specification tweaks to ensure the measure remains up-to-date with current practice and based on feedback received from measure users. There have been no changes to the measure specifications since the last measure update.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

***IF an OUTCOME MEASURE**, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

Women who were tested for chlamydia during the measurement year.

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

***IF an OUTCOME MEASURE,** describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

Women who had at least one test for chlamydia (Chlamydia Tests Value Set) during the measurement year.

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

Women 16-24 years of age who had a claim or encounter indicating sexual activity.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

***IF an OUTCOME MEASURE**, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

Women 16-24 years of age as of December 31 of the measurement year who were identified as sexually active during the measurement year. Two methods are used to identify sexually active women: claim/encounter data and pharmacy data. Both methods are used to identify the eligible population; however, women only need to be identified in one method to be eligible for the measure.

Claim/encounter data: women who had a claim or encounter indicating sexual activity during the measurement year. A code from any of the following meet criteria: Pregnancy Value Set, Sexual Activity Value Set, Pregnancy Tests Value Set.

Pharmacy data: women who were dispensed prescription contraceptives during the measurement year.

Contraceptives Medications List

--Contraceptives: Desogestrel-ethinyl estradiol; Dienogest-estradiol (multiphasic); Drospirenone-ethinyl estradiol; Drospirenone-ethinyl estradiol-levomefolate (biphasic); Ethinyl estradiol-ethynodiol; Ethinyl estradiol-etonogestrel; Ethinyl estradiol-levonorgestrel; Ethinyl estradiol-norelgestromin; Ethinyl estradiol-norethindrone; Ethinyl estradiol-norgestimate; Ethinyl estradiol-norgestrel; Etonogestrel; Levonorgestrel; Medroxyprogesterone; Mestranol-norethindrone; Norethindrone

--Diaphragm

--Spermicide: Nonxynol 9

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

Women who received a pregnancy test to determine contraindications for medication (isotretinoin) or x-ray.

Women who were in hospice or using hospice services during the measurement year.

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

Exclude women who were identified as sexually active based on a pregnancy test alone (Pregnancy Tests Value Set) AND who met either of the following:

1) A pregnancy test (Pregnancy Test Exclusion Value Set) during the measurement year AND a prescription for isotretinoin on the date of the pregnancy test or the 6 days after the pregnancy test.
2) A pregnancy test (Pregnancy Test Exclusion Value Set) during the measurement year AND an x-ray (Diagnostic Radiology Value Set) on the date of the pregnancy test or the 6 days after the pregnancy test.

Retinoid Medications: Isotretinoin

Exclude women who were in hospice or using hospice services during the measurement year.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

The measure includes two age stratifications and a total rate:

1) 16-20 years.
2) 21-24 years.
3) Total

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

Refer to items S.7 (Denominator details) and S.2b (Data Dictionary) for tables.

Step 1. Determine the eligible population. Identify all women 16-24 years of age as of December 31 of the measurement year who were identified as sexually active during the measurement year. Two methods are used to identify sexually active women: pharmacy data (see Contraceptives Medications List) and claim/encounter data (Pregnancy Value Set, Sexual Activity Value Set, and Pregnancy Tests Value Set). Both methods are used to identify the eligible population; however, women only need to be identified in one method to be eligible for the measure.

Step 2. Exclude women who qualified for the eligible population based on a pregnancy test (Pregnancy Tests Value Set) alone AND who meet either of the following: (1) A pregnancy test (Pregnancy Test Exclusion Value Set) during the measurement year AND a prescription for isotretinoin on the date of the pregnancy test or the 6 days after the pregnancy test; or (2) A pregnancy test (Pregnancy Test Exclusion Value Set) during the measurement year AND an x-ray (Diagnostic Radiology Value Set) on the date of the pregnancy test or the 6 days after the pregnancy test. Exclude women who used hospice services or elected to use a hospice benefit any time during the measurement year, regardless of when the services began.

Step 3. Determine the denominator: eligible population minus exclusions.

Step 4. Determine the numerator. Determine the number of women in the denominator who had at least one chlamydia test (Chlamydia Tests Value Set) during the measurement year.

Step 5. Report two age stratifications (16-20 years and 21-24 years), and a total rate. The total is the sum of the age stratifications.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

N/A

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

N/A

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Claims, Enrollment Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g., name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from health plans via NCQA's online data submission system.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

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

Health Plan

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Outpatient Services

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

nqf_testing_attachment_7.1_508Compliant.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

No - This measure is not risk-adjusted

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 0033
Measure Title: Chlamydia Screening in Women (CHL)
Date of Submission: 4/9/2021

Type of Measure:

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

*cell intentionally left blank

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

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

2021 Submission

N/A

2016 Submission

N/A

1.3. What are the dates of the data used in testing?

2021 Submission

HEDIS Health Plan performance data from measurement year 2019.

2016 Submission

HEDIS Health Plan performance data from 2012-2014.

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	Measure Tested at Level of:
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☐ other:	☐ other:

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

2021 Submission

This measure evaluates the percentage of women aged 16-24 enrolled in commercial and Medicaid health plans who were identified as sexually active and who had at least one test for chlamydia during the measurement year. The intended use of the measure is to assess the quality of care provided by health plans for the female adolescent/young adult population. As required by the specified level of accountability, we assessed measure score reliability testing and construct validity testing using data from all health plans reporting the HEDIS measure to NCQA in 2019. These data came from 402 commercial plans and 251 Medicaid plans in total, which were geographically diverse and varied in size.

Systematic evaluation of face validity was assessed during measure development with the independent panel of experts described below in the 2016 submission, as well as several other NCQA panels: the Technical Measurement Advisory Panel (includes 12 members, including representation by health plan methodologists, clinicians, HEDIS auditors and state/federal users of measures) and the Committee on Performance Measurement (CPM), which oversees measures used in NCQA programs and includes representation by purchasers, consumers, health plans, health care providers, and policy makers. This panel is composed of 21 independent members that reflect the diversity of constituencies that performance measurement serves. The CPM's recommendations are reviewed and approved by NCQA's Board of Directors.

2016 Submission

NCQA tested the measure for face validity using a panel of stakeholders with specific expertise in measurement of women and child health care. This panel included representatives from key stakeholder groups, including experts on women's health, family physicians, health plans, AHRQ and other researchers in the field. (See list of members of Women & Child Measurement Advisory Panel (WCMAP). Experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

2021 Submission

HEDIS data are summarized at the health plan level and stratified by plan type (e.g., commercial and Medicaid). Below is a description of the sample. It includes the number of health plans submitting the measure for HEDIS and the median eligible population for the measure across plans, stratified by age.

Table 1. Median eligible population for *Chlamydia Screening in Women* by age and plan type, 2019

Plan	Number of Plans	Age	Median number of eligible women per plan
Commercial	393	16-20 years	1,385
Commercial	396	21-24 years	1,665
Commercial	402	Total	3,012
Medicaid	247	16-20 years	2,261
Medicaid	236	21-24 years	1,482
Medicaid	251	Total	3,796

2016 Submission

This measure assesses the percentage women 16-24 years of age who were identified as sexually active and who received a test for chlamydia. The improvement in quality envisioned by the use of this measure is increased identification of untreated chlamydia infections in women that can lead to serious and irreversible complications and can be unknowingly transmitted to sexual partners. Despite the availability of effective treatments, a large proportion of sexually active individuals continue to go undiagnosed due to the disease's asymptomatic nature. Early detection, screening, and treatment have proven to be effective in managing and preventing chlamydia.

The data references are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. In 2013, HEDIS measures covered more than 171 million people from 814 HMOs and 353 PPOs. Below is a description of the denominator for this measure. It includes the number of health plans included in HEDIS data collection and the median eligible population for the measure across health plans.

Commercial—16 to 20 years of age

YEAR	N Plans	Median Denominator Size per Plan
2014	401	1,323
2013	409	1,390
2012	411	1,376

Commercial—21 to 24 years of age

YEAR	N Plans	Median Denominator Size per Plan
2014	403	1,561
2013	410	1,590
2012	411	1,549

Commercial—Total

YEAR	N Plans	Median Denominator Size per Plan
2014	405	2,922
2013	412	2,984
2012	415	2,956

Medicaid—6 to 20 years of age

YEAR	N Plans	Median Denominator Size per Plan
2014	189	1,979
2013	191	1,556
2012	171	1,655

Medicaid—21 to 24 years of age

YEAR	N Plans	Median Denominator Size per Plan
2014	195	1,237
2013	190	957
2012	172	1,034

Medicaid—Total

YEAR	N Plans	Median Denominator Size per Plan
2014	198	3,082
2013	198	2,362
2012	176	2,508

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

2021 Submission

There are no differences in the data used for reliability and construct validity testing. The systematic assessment of face validity was done with multi-stakeholder experts as described in Section 1.5 above.

1.8 What were 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.

2021 Submission

Social risk factor data were not available in reported results. This measure is specified to be reported separately by Medicaid and commercial plan types, which serves as a proxy for income and other socioeconomic factors. NCQA is actively engaged with partners including the CMS Office of Minority Health in identifying feasible methods to further integrate social risk factors into health plan quality measures, with a focus on stratification. This is aligned with recent recommendations from MedPAC and ASPE on optimal methods for addressing social risk in quality measurement and programs.[1,2]This is an NCQA wide initiative. Our intent is to implement methods to bridge data concerns in the future.

1. Medicare Payment Advisory Commission. (2020). The Medicare Advantage program: Status report. In Report to the Congress: Medicare Payment Policy (p. 397). http://medpac.gov/docs/default-source/reports/mar20_medpac_ch13_sec.pdf
2. Office of the Assistant Secretary for Planning and Evaluation, & U.S. Department of Health & Human Services. (2020). Second Report to Congress on Social Risk and Medicare's Value-Based Purchasing Programs. https://aspe.hhs.gov/social-risk-factors-and-medicares-value-basedpurchasing-programs

2016 Submission

HEDIS data, including data for this measure, are stratified by type of insurance (i.e., Commercial, Medicaid, Medicare). NCQA does not currently collect performance data stratified by race, ethnicity, or language. Escarce et al. have described in detail the difficulty of collecting valid data on race, ethnicity and language at the health plan level (Escarce 2011). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. The HEDIS Health Plan Measure Set contains two measures that can assist with stratification to assess health care disparities. The Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing and using race/ethnicity and language data to assess health care disparities. Based on extensive work by NCQA to understand how to promote culturally and linguistically appropriate services among plans and providers, we have many examples of how health plans have used HEDIS measures to design quality improvement programs to decrease disparities in care.

Escarce, J.J., R. Carreon, G. Vesolovskiy, E.H. Lawson. 2011. "Collection Of Race And Ethnicity Data By Health Plans Has Grown Substantially, But Opportunities Remain To Expand Efforts." Health Affairs 30(10): 1984-1991. doi: 10.1377/hlthaff.2010.1117.

2a2. RELIABILITY TESTING

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

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

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

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

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

2021 Submission

We utilized the methodology described by John Adams (Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009) to calculate signal-to-noise reliability. This methodology uses the Beta-binomial model to assess how well one can confidently distinguish the performance of one reporting entity from another. Conceptually, the Beta-binomial model is the ratio of signal to noise. The signal is the proportion of the variability in measured performance that can be explained by real differences across reporting entities (plans, physicians, etc.) in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures, such as the *Chlamydia Screening in Women* measure. Reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (i.e., noise), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance across reporting entities.

For the *Chlamydia Screening in Women* measure, submissions (plans) are the reporting entity. For the formulas and explanations below, we use submissions (plans) as the reporting entity.

The formula for signal-to-noise reliability is:

$$\text{Signal-to-noise reliability} = \sigma^2_{\text{plan-to-plan}} / (\sigma^2_{\text{plan-to-plan}} + \sigma^2_{\text{error}})$$

Therefore, we need to estimate two variances: 1) variance between plans ($\sigma^2_{\text{plan-to-plan}}$); 2) variance within plans (σ^2_{error}).

1. Variance between plans = $\sigma^2_{\text{plan-to-plan}} = (\alpha\ \beta) / (\alpha + \beta + 1)(\alpha + \beta)^2$

   α and β are two shape parameters of the Beta-Binomial distribution, $\alpha > 0$, $\beta > 0$

2. Variance within plans: $\sigma^2_{\text{error}} = \hat{p}(1- \hat{p})/n$

   $\hat{p}$ = observed rate for the plan

   n = plan-specific denominator for the observed rate (most often, including for CHL, n is the number of eligible plan members)

Using Adams' 2009 methodology, we estimated the reliability for each reporting entity, then averaged these reliability estimates across all reporting entities to produce a point estimate of signal-to-noise reliability. We label this point estimate "mean signal-to-noise reliability". The mean signal-to-noise reliability measures how well, on average, the measure can differentiate between reporting entity performance on the measure.

Along with the point estimate of mean signal-to-noise reliability, we are also providing:

1. **The standard error (SE) and 95% confidence interval (95% CI) of the mean signal-to-noise reliability for all plans and stratified by the denominator size (number of eligible members per plan)**. The SE and 95% CI of the mean signal-to-noise reliability provides information about the stability of reliability. The 95% CI is the mean signal-to-noise reliability ± (1.96*SE). The narrower the confidence interval, the less the mean signal-to-noise reliability estimate will change due to idiosyncratic features of specific plans. We also stratified the results by the denominator size using terciles of the distribution to provide additional information about the stability of reliability.

2. **The distribution (minimum, 10th, 25th, 50th, 75th, 90th, maximum) of the plan-level signal-to-noise reliability estimates**. Each plan's reliability estimate is a ratio of signal to noise, as described above [$\sigma^2_{\text{plan-to-plan}} / (\sigma^2_{\text{plan-to-plan}} + \sigma^2_{\text{error}})$]. Variability between plans ($\sigma^2_{\text{plan-to-plan}}$) is the same for each plan, while the specific plan error (σ^2_{error}) varies. Reliability for each plan is an ordinal measure of how well one can determine where a plan lies in the distribution across plans, with higher estimates indicating better reliability. We also stratified the results by the denominator size using terciles of the

distribution to provide additional information about the distribution of plan-level signal-to-noise reliability estimates. The number of plans in each stratum and the per-plan denominators of the performance rates are displayed in the summary tables.

This methodology allows us to estimate the reliability for each plan and summarize the distribution of these estimates.

2016 Submission

Reliability was estimated by using the beta-binomial model. Beta-binomial is a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® health plan measures. The beta-binomial model assumes the plan score is a binomial random variable conditional on the plan´s true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates. The beta distribution can be symmetric, skewed or even U-shaped.

Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

2021 Submission

Signal-to-Noise Reliability Assessment for *Chlamydia Screening in Women* Measure

Table 2a. Point estimates of mean signal-to-noise reliability using above methodology.

Chlamydia Screening in Women	Point Estimate: Mean Signal-to-Noise Reliability (Commercial)	Point Estimate: Mean Signal-to-Noise Reliability (Medicaid)
Chlamydia Screening in Women (16-20)	0.975	0.978
Chlamydia Screening in Women (21-24)	0.964	0.955
Chlamydia Screening in Women (Total)	0.979	0.984

Table 2b. Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) for the *Chlamydia Screening in Women (Total Ages 16-24)* Measure by Terciles of the Denominator Size and for All Submissions Stratified by Plan Type, Calendar Year 2019 Data

Stratification	Number of Plans	Number of Eligible Members per Plan (min–max)	Mean Signal-To-Noise Reliability	SE	95% CI
All Commercial	**402**	**31–128,323**	**0.979**	**0.002**	**(0.975, 0.984)**
Tercile 1	133	31–1,281	0.941	0.0064	(0.928, 0.953)
Tercile 2	132	1313–6,085	0.994	0.0002	(0.993, 0.994)
Tercile 3	137	6121–128,323	0.998	0.0001	(0.998, 0.999)
All Medicaid	**251**	**71–61,541**	**0.983**	**0.002**	**(0.980, 0.988)**
Tercile 1	83	71–1,726	0.969	0.0037	(0.962, 0.976)
Tercile 2	83	1,744–6,452	0.994	0.0003	(0.994, 0.995)
Tercile 3	85	6,465–61,541	0.997	0.0001	(0.997, 0.997)

SE: Standard Error of the mean. 95% CI: 95% confidence interval.

Table 3. Distribution of Plan-Level Signal-To-Noise Reliability for the *Chlamydia Screening in Women (Total Ages 16-24)* measure by Terciles of the Denominator Size and for All Submissions by Plan Type, Calendar Year 2019 Data

Stratification	Number of Plans	Distribution of Plan Estimates of Signal-to-Noise Reliability: Min	Distribution of Plan Estimates of Signal-to-Noise Reliability: P10	Distribution of Plan Estimates of Signal-to-Noise Reliability: P25	Distribution of Plan Estimates of Signal-to-Noise Reliability: P50	Distribution of Plan Estimates of Signal-to-Noise Reliability: P75	Distribution of Plan Estimates of Signal-to-Noise Reliability: P90	Distribution of Plan Estimates of Signal-to-Noise Reliability: Max
All Commercial	**402**	0.657	0.959	0.980	0.994	0.998	0.999	1.000
Tercile 1	133	0.632	0.865	0.949	0.968	0.978	0.982	0.985
Tercile 2	133	0.987	0.989	0.992	0.994	0.996	0.997	0.998
Tercile 3	136	0.997	0.997	0.998	0.999	0.999	1.000	1.000
All Medicaid	**251**	0.790	0.957	0.985	0.996	0.998	0.999	1.000
Tercile 1	83	0.843	0.927	0.963	0.983	0.990	0.992	0.994
Tercile 2	83	0.989	0.991	0.992	0.995	0.996	0.997	0.998
Tercile 3	85	0.994	0.995	0.996	0.997	0.998	0.999	1.000

2016 Submission

Reliability statistic for Chlamydia screening is 0.99.

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

2021 Submission

In general, a score of 0.7 or higher suggests the measure has adequate reliability. The results suggest the measure has high reliability and more details are discussed below.

Table 2b provides the point estimate of mean signal-to-noise reliability, its standard error, and the 95% CI for the *Chlamydia Screening in Women (Total Ages 16-24)* measure for commercial and Medicaid plans overall and stratified by the denominator size (distribution of the number of eligible members per plan). Across all commercial plans, the reliability estimate is 0.979, and the 95% CI is (0.975, 0.984), indicating very good reliability. Stratified analyses show that reliability increase as plan size gets larger and stay above 0.9. Across all Medicaid plans, the reliability estimate is 0.984 and the 95% CI is (0.980, 0.988), indicating very good reliability. Results from the stratified analyses show that reliability exceeds 0.9 for all terciles.

Table 3 summarizes the distribution of plan-level signal-to-noise reliability estimates for the *Chlamydia Screening in Women (Total Ages 16-24)* measure. Across all commercial plans, the estimates range from 0.657 to 1.0. The 50th percentile is 0.994, which exceeds the 0.70 threshold for reliability. For Medicaid plans, the estimates range from 0.790 to 1.0; the 10th percentile is 0.96, indicating very good reliability. This table also include the distribution of plan-level signal-to-noise reliability estimates stratified by denominator size. Reliability estimates are higher for plans with a larger denominator.

2016 Submission

This question was not on the 2016 form.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

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

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*
2021 Submission

Method of testing construct validity

We tested for construct validity by exploring the following:

- Is the *Chlamydia Screening in Women* measure positively correlated with the *Cervical Cancer Screening* measure, which assesses the percentage of women 21-64 years of age who were screened for cervical cancer using appropriate methods?

NCQA performs Pearson correlation for construct validity using HEDIS health plan data. The test estimates the strength of linear association between two continuous variables: the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a strong positive linear association: an increase in values of one variable is associated with increase in value of another variable. A value of 0 indicates no linear association. A value of -1 indicates a strong negative relationship in which an increase in values of the first variable is associated with a decrease in values of the second variable. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone.

Method of assessing face validity

NCQA develops measures using a standardized process. For new measures, face validity is assessed at various steps as described below.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical measurement advisory panels (MAPs), whose members are authorities on clinical priorities for measurement, participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness, and feasibility. This information is gathered into a work-up format, which is vetted by the MAPs, including the Behavioral Health Measurement Advisory Panel (BHMAP), Geriatric Measurement Advisory Panel (GMAP), the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM) as well as other panels as necessary.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. At this step, face validity is systematically determined by the CPM, which uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA about proposed new measures. Public comment offers an opportunity to assess the validity, feasibility, importance, and other attributes of a measure from a wider audience. For this measure, a majority of public comment respondents supported the measure. NCQA MAPs and the technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. Face validity is then again systematically assessed by the CPM. The CPM reviews all comments before making a final decision and votes to recommend approval of new measures for HEDIS. NCQA's Board of Directors then approves new measures.

ICD-10 CONVERSION:

Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

Steps in ICD-9 to ICD-10 Conversion Process

1. NCQA staff identify ICD-10 codes to be considered based on ICD-9 codes currently in measure. Use GEM to identify ICD-10 codes that map to ICD-9 codes. Review GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA staff identify additional codes (not identified by GEM mapping step) that should be considered. Using ICD-10 tabular list and ICD-10 Index, search by diagnosis or procedure name for appropriate codes.

3. NCQA HEDIS Expert Coding Panel review NCQA staff recommendations and provide feedback.
4. As needed, NCQA Measurement Advisory Panels perform clinical review. Due to increased specificity in ICD-10, new codes and definitions require review to confirm the diagnosis or procedure is intended to be included in the scope of the measure. Not all ICD-10 recommendations are reviewed by NCQA MAP; MAP review items are identified during staff conversion or by HEDIS Expert Coding Panel.
5. Post ICD-10 code recommendations for public review and comment.
6. Reconcile public comments. Obtain additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
7. NCQA staff finalize ICD-10 code recommendations.

Tools Used to Identify/Map to ICD-10

All tools used for mapping/code identification from CMS ICD-10 website

(http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html

GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

Expert Participation

The NCQA HEDIS Expert Coding Panel and NCQA's Diabetes Expert Panel reviewed and provided feedback on staff recommendations. Names and credentials of the experts who served on these panels are listed under **Additional Information, Ad.1. Workgroup/Expert Panel Involved in Measure Development.**

Summary of Stakeholder Comments Received

NCQA posted ICD-10 codes for public review and comment in March 2011 and March 2012. NCQA received comments from four organizations:

- Support recommendations
- Questions about select codes
- Recommended additional codes for consideration

2016 Submission

NCQA tested the measure for face validity using a panel of stakeholders with specific expertise in measurement of women and child health care. This panel included representatives from key stakeholder groups, including experts on women's health, family physicians, health plans, AHRQ and other researchers in the field. (See list of members of Women & Child Measurement Advisory Panel (WCMAP)). Experts reviewed the results of the field test and assessed whether the results were consistent with expectations, whether the measure represented quality care, and whether we were measuring the most important aspect of care in this area.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

2021 Submission

Table 4. Results of Pearson Correlation Coefficient for Commercial and Medicaid health plans for the *Chlamydia Screening in Women (Total Ages 16-24)* Measure, 2019

Commercial vs Medicaid	Cervical Cancer Screening
Commercial	0.53
(N=, p value =)	(402, p < 0.001)
Medicaid	0.32
(N=, p value =)	(238, p < 0.001)

Table 5. Results of Pearson Correlation Coefficient for Commercial and Medicaid health plans for the *Chlamydia Screening in Women (Ages 21–24)* Measure, 2019

Commercial vs Medicaid	Cervical Cancer Screening
Commercial	0.53
(N=, p value =)	(396, p < 0.001)
Medicaid	0.44
(N=, p value =)	(236, p < 0.001)

2016 Submission

This measure was deemed valid by the expert panel.

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

2021 Submission

Interpretation of construct validity testing

For the purposes of this analysis and the intended use of this measure to evaluate the quality of care for members across health plans, correlation was considered high (strong) if the correlation coefficient is 0.75 to 1, moderate if 0.25 to 0.75, and low (weak) if 0 to 0.25.

Chlamydia Screening in Women (Total Ages 16-24) is positively correlated with *Cervical Cancer Screening* in both commercial (correlation coefficient = 0.53, p < 0.001) and Medicaid (correlation coefficient = 0.32, p <0.001) product lines. The correlation is significant, indicating both preventive screening measures for women are positively associated, a better performance rate in cervical cancer screening is associated with better performance rate in Chlamydia screening. The results indicate that the measure has moderate validity.

The difference in age ranges between the two measures may have impacted the correlation coefficient.

Cervical cancer screening is recommended and specified for a wider age group (21–64 years of age) than chlamydia screening (16–24 years of age). If we look at the correlation between *Chlamydia Screening in Women* in the upper age range of the measure (ages 21–24) and *Cervical Cancer Screening*, the coefficients are higher than the correlations between *Chlamydia Screening in Women Ages 16-24* and *Cervical Cancer Screening*. This underlying population difference may explain why the correlation coefficients are moderate, rather than high.

Interpretation of systematic assessment of face validity

The multi-stakeholder advisory panels concluded the measures had good face validity.

2016 Submission

This question was not on the 2016 form.

_____

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b3*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2021 Submission:

NCQA currently allows health plans to apply exclusions to their results. NCQA does not collect data on exclusions for HEDIS reporting of the measure. In measure development and field testing, we investigated and validated the exclusion applied to the eligible denominator.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2021 Submission:

N/A

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. **Note: If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

2021 Submission:

Experts on our measurement advisory panels recommended specifying the exclusions in the measure based on the clinical rationale and from an accountability perspective, and because it is feasible to collect the data with minimal burden.

_____

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2021 Submission

N/A. Not an intermediate or health outcome, PRO-PM, or resource use measure.

2016 Submission

Risk adjustment is not applied for this measure at the health plan level. NCQA has determined that risk adjustment is not necessary other than the reporting of the measure is stratified by insurance coverage (commercial and Medicaid). The measure is stratified by age and product line.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g., prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

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

***If stratified, skip to* 2b3.9**

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b3.8. Statistical Risk Model Calibration–Risk decile plots or calibration curves:

2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b3.11. Optional Additional Testing for Risk Adjustment (***not required****, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

2021 Submission

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure.

To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the below 25th and above 75th percentile groups. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared against a normal distribution. If the p-value of the test statistic is less than .05, then the two plans' performance is significantly different from each other.

2016 Submission

Comparison of means and percentiles; analysis of variance against established benchmarks; if sample size is >400, we would use an analysis of variance.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

2021 Submission

Table 6. Variation in Performance for Commercial and Medicaid health plans for the *Chlamydia Screening in Women* Measure, 2019

Plan Type	Rate	N	Avg. EP	Avg.	SD	P10	P25	P50	P75	P90	IQR	p-value
Commercial	Age 16-20	393	4,176	44%	13%	30%	35%	42%	51%	63%	16%	< 0.001
Commercial	Age 21-24	396	4,816	55%	10%	43%	48%	54%	61%	70%	13%	< 0.001
Commercial	Total	402	8,827	50%	11%	36%	42%	48%	56%	66%	14%	< 0.001
Medicaid	Age 16-20	247	4,036	55%	13%	39%	47%	54%	63%	70%	16%	< 0.001
Medicaid	Age 21-24	236	2,780	64%	10%	55%	60%	65%	70%	74%	10%	< 0.001
Medicaid	Total	251	6,586	58%	12%	43%	51%	58%	66%	71%	15%	< 0.001

N = Number of plans reporting

IQR = Interquartile range

p-value = p-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile.

2016 Submission

The following data are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. Performance data are summarized at the health plan level by mean, standard deviation, and performance at the 10th, 25th, 50th, 75th and 90th percentile. Data are stratified by year and product line (i.e., commercial and Medicaid).

Chlamydia Screening in Women

Commercial Rate—16 to 20 years of age

YEAR	MEAN	ST DEV	10TH	25TH	50TH	75TH	90TH	Interquartile Range
2014	40%	11%	29%	33%	38%	45%	56%	12
2013	40%	10%	29%	33%	38%	45%	54%	12
2012	40%	10%	30%	34%	38%	45%	53%	12

Commercial Rate—21 to 24 years of age

YEAR	MEAN	ST DEV	10TH	25TH	50TH	75TH	90TH	Interquartile Range
2014	49%	10%	38%	43%	47%	56%	63%	13
2013	48%	10%	36%	41%	46%	55%	63%	14
2012	47%	10%	34%	40%	46%	54%	62%	14

Commercial Rate—Total

YEAR	MEAN	ST DEV	10TH	25TH	50TH	75TH	90TH	Interquartile Range
2014	45%	10%	34%	38%	43%	51%	60%	13
2013	44%	10%	33%	37%	43%	51%	58%	13
2012	44%	10%	32%	37%	42%	50%	57%	13

Medicaid Rate—16 to 20 years of age

YEAR	MEAN	ST DEV	10TH	25TH	50TH	75TH	90TH	Interquartile Range
2014	51%	11%	37%	44%	50%	58%	67%	13
2013	51%	11%	37%	44%	52%	59%	64%	15
2012	53%	10%	41%	47%	54%	59%	66%	13

Medicaid Rate—21 to 24 years of age

YEAR	MEAN	ST DEV	10TH	25TH	50TH	75TH	90TH	Interquartile Range
2014	60%	10%	47%	54%	61%	67%	72%	13
2013	62%	10%	49%	57%	63%	69%	72%	13
2012	63%	9%	52%	59%	65%	71%	73%	12

Medicaid Rate—Total

YEAR	MEAN	ST DEV	10TH	25TH	50TH	75TH	90TH	Interquartile Range
2014	55%	11%	40%	49%	54%	62%	69%	13
2013	55%	10%	41%	49%	55%	63%	67%	14
2012	57%	10%	46%	51%	57%	64%	69%	13

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

2021 Submission

Table 6 summarizes the distribution of plan-level performance for the *Chlamydia Screening in Women* measure for commercial and Medicaid Plans. For the *Total rate (ages 16-24),* there is a 0.14 gap in performance between commercial plans at the 25th and 75th percentiles, and a 0.15 gap in performance among Medicaid plans. The difference in performance between plans in the 25th percentile and 75th percentile is statistically significant.

For commercial health plans, the large gap in performance between 25th and 75th percentile plans represents an average 1,236 more women ages 16-24 receiving chlamydia screening in high performing commercial plans compared to low performing commercial plans (estimated from average health plan eligible population). For Medicaid health plans, the large gap in performance between 25th and 75th percentile plans represents an average 988 more women ages 16-24 receiving chlamydia screening in high performing Medicaid plans compared to low performing Medicaid plans (estimated from average health plan eligible population).

2016 Submission

This question was not on the 2016 form.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

Note*: This item is directed to measures that are risk-adjusted (with or without 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 eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without 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.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

2021 Submission

HEDIS measures apply to enrolled members in a health plan, and NCQA has a rigorous audit process to ensure the eligible population and numerator events for each measure are correctly identified and reported. The audit process is designed to verify primary data sources used to populate measures and ensure specifications are correctly implemented.

The HEDIS Compliance Audit addresses the following functions:

- Information practices and control procedures
- Sampling methods and procedures
- Data integrity
- Compliance with HEDIS specifications
- Analytic file production
- Reporting and documentation

2016 Submission

This question was not on the 2016 form.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; **if no empirical sensitivity analysis**, identify the approaches for handling missing data that were considered and pros and cons of each*)

2021 Submission

HEDIS addresses missing data in a structured way through its audit process. HEDIS measures apply to enrolled members in a health plan, and NCQA-certified auditors use standard audit methodologies to assess whether data sources are missing data. If a data source is found to be missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used. Once measures are added to HEDIS, NCQA conducts a first-year analysis to assess the feasibility of the measure when widely implemented in the field. This analysis includes an assessment of how many plans report valid rates vs. rates that are materially biased (or have other issues, such as small denominators). These considerations are weighed in the deliberation process before measures are approved for public reporting.

2016 Submission

This question was not on the 2016 form.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias? (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; **if no empirical analysis**, provide rationale for the selected approach for missing data*)

2021 Submission

All health plans that reported 2019 HEDIS data for this measure reported valid rates as determined by NCQA-certified auditors. This means that auditors did not find any missing data sources for any of the health plan data submissions and determined that none of the rates were materially biased.

2016 Submission

This question was not on the 2016 form.

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. Data Elements Generated as Byproduct of Care Processes.

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

N/A

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

Attachment:

3c. Data Collection Strategy

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

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO's ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable apples-to-apples comparisons between health plans.

The HEDIS Compliance Audit addresses the following functions:

1) Information practices and control procedures
2) Sampling methods and procedures
3) Data integrity
4) Compliance with HEDIS specifications
5) Analytic file production
6) Reporting and documentation

In addition to the HEDIS audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system, NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measures. This system is vital to the regular re-evaluation of the NCQA measures.

Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

Broad public use and dissemination of these measures are encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Plan for	Public Reporting CMS Medicaid Adult Core Set https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-core-set/index.html CMS Qualified Health Plan (QHP) Quality Rating System (QRS) https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/2017_QRS_and_QHP_Enrollee_Survey_Technical_Guidance.pdf NCQA Health Plan Rating/Report Card http://reportcard.ncqa.org/plan/external/plansearch.aspx CMS Medicaid Adult Core Set https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-core-set/index.html CMS Qualified Health Plan (QHP) Quality Rating System (QRS) https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/2017_QRS_and_QHP_Enrollee_Survey_Technical_Guidance.pdf NCQA Health Plan Rating/Report Card http://reportcard.ncqa.org/plan/external/plansearch.aspx Payment Program California Align. Measure. Perform. (AMP) Commercial HMO Program https://www.iha.org/performance-measurement/amp-program/amp-program-descriptions/ California Align. Measure. Perform. (AMP) Medi-Cal Managed Care Program https://www.iha.org/performance-measurement/amp-program/amp-program-descriptions/ California Align. Measure. Perform. (AMP) Commercial HMO Program https://www.iha.org/performance-measurement/amp-program/amp-program-descriptions/ California Align. Measure. Perform. (AMP) Medi-Cal Managed Care Program https://www.iha.org/performance-measurement/amp-program/amp-program-descriptions/ Regulatory and Accreditation Programs NCQA Health Plan Accreditation http://www.ncqa.org/tabid/123/Default.aspx NCQA Accountable Care Organization Accreditation http://www.ncqa.org/Programs/OtherPrograms/acomeasuresPilotProject.aspx NCQA Health Plan Accreditation http://www.ncqa.org/tabid/123/Default.aspx NCQA Accountable Care Organization Accreditation http://www.ncqa.org/Programs/OtherPrograms/acomeasuresPilotProject.aspx Quality Improvement (external benchmarking to organizations) NCQA Quality Compass http://www.ncqa.org/tabid/177/Default.aspx NCQA Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

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

CALIFORNIA ALIGN. MEASURE. PERFORM. (AMP) COMMERCIAL HMO PROGRAM: This measure is used in California's (AMP) Commercial HMO program. California's AMP programs focus on creating comprehensive benchmarks and a reliable assessment of performance for medical groups, independent practice association (IPAs), and accountable care organizations (ACOs) across health plans. The AMP Commercial HMO program (formerly known as Value Based Pay for Performance) is the cornerstone upon which all of IHA's performance measurement programs were built. Initiated in 2001, the program now includes participation from eleven health plans and about 200 California physician organizations caring for over 9 million Californians enrolled in commercial HMO and point of service products—
representing 95% of commercial HMO enrollment in the state. AMP Commercial HMO has four key components: a common set of measures and benchmarks that spans clinical quality, patient experience, utilization, and cost of care measures; value-based health plan incentive payments to physician organizations; public reporting of Triple Aim performance results for physician organizations; and public recognition awards.
CALIFORNIA ALIGN. MEASURE. PERFORM. (AMP) MEDI-CAL MANAGED CARE PROGRAM: This measure is used in California's (AMP) Medi-Cal Managed Care program. California's AMP programs focus on creating comprehensive benchmarks and a reliable assessment of performance for medical groups, independent practice association (IPAs), and accountable care organizations (ACOs) across health plans. The AMP Medi-Cal Managed Care program is based on a common set of measures and benchmarks that spans clinical quality, patient experience, utilization, and cost of care measures. The program collects data and calculates performance results for medical groups, IPAs and FQHCs that provide care to Medi-Cal Managed Care enrollees. Health plans can use the results to make value-based incentive payments to their contracted providers.
MEDICAID ADULT CORE SET: There are a core set of health quality measures for Medicaid-enrolled adults. The Medicaid Adult Core Set was identified by the Centers of Medicare & Medicaid (CMS) in partnership with the Agency for Healthcare Research and Quality (AHRQ). The data collected from these measures will help CMS to better understand the quality of health care that adults enrolled in Medicaid receive nationally. Beginning in January 2014 and every three years thereafter, the Secretary is required to report to Congress on the quality of care received by adults enrolled in Medicaid. Additionally, as of 2014, state data on the adult quality measures is part of the Secretary's annual report on the quality of care for adults enrolled in Medicaid.
NCQA HEALTH PLAN RATING/REPORT CARDS: This measure is used to calculate health plan rankings which are reported on the NCQA website. These rankings are based on performance on HEDIS measures among other factors. In 2019, a total of 515 commercial health plans and 188 Medicaid health plans across 50 states were included in the rankings.
NCQA STATE OF HEALTH CARE ANNUAL REPORT: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2019, the report included results from calendar year 2018 for health plans covering a record 136 million people, or 43 percent of the U.S. population.
NCQA HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of commercial and Medicaid health plans. In 2019, 336 commercial health plans covering 87 million lives and 77 Medicaid health plans covering 9.1 million lives were accredited. Health plans are scored based on performance compared to benchmarks.
NCQA ACCOUNTABLE CARE ORGANIZATION ACCREDITATION: This measure is used in NCQA's ACO Accreditation program, that helps health care organizations demonstrate their ability to improve quality, reduce costs and coordinate patient care. ACO standards and guidelines incorporate whole-person care coordination throughout the health care system.

NCQA QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.
QUALIFIED HEALTH PLAN (QHP) QUALITY RATING SYSTEM (QRS): This measure is used in the Qualified Health Plan (QHP) Quality Rating System, which provides comparable information to consumers about the quality of health care services and QHP enrollee experience offered in the Marketplaces.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

N/A

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

Health plans that report HEDIS calculate their rates and know their performance when submitting to NCQA. NCQA publicly reports rates across all plans and also creates benchmarks in order to help plans understand how they perform relative to other plans. Public reporting and benchmarking are effective quality improvement methods.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

NCQA publishes HEDIS results annually in our Quality Compass tool. NCQA also presents data at various conferences and webinars. For example, at the annual HEDIS Quality Congress, NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly. NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System, as described in Section 3c.1.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

NCQA measures are evaluated regularly using a consensus-based process to consider input from multiple stakeholders, including but not limited to entities being measured. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System. This information enables NCQA to comprehensively assess a measure's adherence to the HEDIS Desirable Attributes of Relevance, Scientific Soundness and Feasibility.

4a2.2.2. Summarize the feedback obtained from those being measured.

Questions received through the Policy Clarification Support system have generally centered around clarification on the definition of "sexually active." Other questions have sought clarification about whether direct optical observation would count as screening.

4a2.2.3. Summarize the feedback obtained from other users

This measure has been deemed a priority measure by NCQA and other entities, as illustrated by its use in programs such as NCQA's Health Plan Accreditation and CMS's Medicaid Adult Core Set program.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

During the measure's regular maintenance cycle, feedback obtained through the mechanisms described in 4a2.2.1 informed how we implemented minor updates to the measure.

Improvement

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

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Over the past three years, this measure has shown slight improvement (approximately 1% improvement over the past three years) across commercial health plans (see section 1b.2 for summary of data from health plans) and consistent average performance of 58% across Medicaid plans. The greatest improvement in performance has been seen for commercial plans (avg. 24% improvement for plans at the minimum performance rate). There is also variation in performance rates when comparing across low- and high-performance plans. For example, in 2019, the percentage point difference between commercial and Medicaid plans in the 10th and 90th percentile was 30 and 28 percentage points, respectively. These gaps indicate a continued opportunity for improvement. These data are nationally representative.

4b2. Unintended Consequences

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

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

There were no identified unintended findings for this measure during testing or since implementation.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

There were no identified unintended benefits for this measure during testing or since implementation.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

Yes

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

NQF #0409 assesses the percentage of patients aged 13 years and older with a diagnosis of HIV/AIDS, who have received chlamydia, gonorrhea, and syphilis screenings at least once since the diagnosis of HIV infection. The measures differ in level of accountability and population of focus. Measure #0409 is a physician level measure and therefore, only includes patients who had an office visit with an eligible provider while NQF #0033 is reported by health plans and includes the entire health plan population. NQF #0409 focuses specifically on patients (both male and female) aged 13 and older that have been diagnosed with HIV/AIDS. Measure 0033 focuses on sexually active female adolescents and young adults, which is aligned to the U.S. Preventive Services Task Force recommendation. In addition, measure 0409 measures screenings at least once since the diagnosis of HIV, while 0033 assesses yearly screening of chlamydia. IMPACT ON INTERPRETABILITY AND DATA COLLECTION BURDEN: The measure performance rates should not be compared, as they focus on different populations of interest.

5b. Competing Measures

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

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A

Appendix

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

No appendix **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): National Committee for Quality Assurance

Co.2 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Co.3 Measure Developer if different from Measure Steward: National Committee for Quality Assurance

Co.4 Point of Contact: Brittany, Wade, wade@ncqa.org, 202-530-0463-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

COMMITTEE ON PERFORMANCE MEASUREMENT

Andrew Baskin, MD CVS Health/Aetna

Elizabeth Drye, MD, SM Yale School of Medicine

Andrea Gelzer, MD, MS, FACP AmeriHealth Caritas

Kate Goodrich, MD, MHS Centers for Medicare & Medicaid Services

David Grossman, MD, MPH Washington Permanente Medical Group

Christine S. Hunter, MD (Co-Chair) Independent Board Director

David K. Kelley, MD, MPA Pennsylvania Department of Human Services

Jeff Kelman, MD, MMSc Department of Health and Human Services

Nancy Lane, PhD Independent Consultant

Bernadette Loftus, MD Independent Consultant

Adrienne Mims, MD, MPH, AGSF, FAAFP Alliant Health Solutions

Amanda Parsons, MD, MBA MetroPlus

Wayne Rawlins, MD, MBA ConnectiCare

Misty Roberts, MSN, RN, CPHQ, PMP Humana

Rodolfo Saenz, MD, MMM, FACOG Riverside Medical Clinic

Marcus Thygeson, MD, MPH (Co-Chair) Bind Benefits

JoAnn Volk, MA Georgetown University Liaisons

Rose Baez, RN, MSN, MBA, CPHQ Blue Cross Blue Shield Association

Jeff Brady, MD, MPH Agency for Healthcare Research and Quality

Ron Kline, MD Office of Personnel Management

Elisa Munthali, MPH National Quality Forum

Chinwe Nwosu, MS America's Health Insurance Plans

Chesley Richards, MD, MPH, FACP Centers for Disease Control and Prevention

Anecia Suneja, CNS-BC Veteran's Health Administration

HEDIS Expert Coding Panel

Glen Braden, MBA, CHCA, Attest Health Care Advisors, LLC

Denene Harper, RHIA, American Hospital Association

DeHandro Hayden, BS, American Medical Association

Patience Hoag, RHIT, CPHQ, CHCA, CCS, CCS-P, Health Services Advisory Group

Nelly Leon-Chisen, RHIA, American Hospital Association

Tammy Marshall, LVN, Aetna

Alec McLure, RHIA, CCS-P, Verisk Health

Michele Mouradian, RN, BSN, McKesson Health Solutions

Craig Thacker, RN, CIGNA HealthCare

Mary Jane F. Toomey, RN CPC, Aetna Better Health

NCINQ Measurement Advisory Panel

Mary Applegate, MD, Ohio Department of Job and Family Services

Katie Brookler, Colorado Department of Health Care Policy and Financing

Cathy Caldwell, MPH, Alabama Department of Public Health

Ted Ganiats, MD, University of California, San Diego

Darcy Gruttadaro, JD, National Allegiance on Mental Illness

Jennifer Havens, MD, NYU School of Medicine

Virginia Moyer, MD, MPH, FAAP, Baylor College of Medicine, USPSTF

Edward Schor, MD, Lucile Packard Foundation for Children's Health

Xavier Sevilla, MD, FAAP, Whole Child Pediatrics

Gwen Smith, Illinois Department of Healthcare and Family Services/Health Management Associates

Janet (Jessie) Sullivan, MD, Hudson Health Plan

Kalahn Taylor-Clark, PhD, MPH, George Mason University

Craig Thiele, MD, CareSource

Jeb Weisman, PhD, Children's Health Fund

Charles Wibbelsman, MD, Kaiser Permanente Medical Group, Inc.

NCINQ Clinician Advisory Panel

Elizabeth Alderman, MD, FAAP, Albert Einstein College of Medicine

Sarah Brewington, MD, Sandhills Pediatrics Inc

Gale Burstein, MD, MPH, FAAP, FSAHM, Women and Children's Hospital of Buffalo, NY

Barry Bzostek, MD, FAAP, Women and Children's Hospital of Buffalo, NY

Danielle Casher, MD, FAAP, St. Christoper's Hospital for Children

Edward Curry, MD, FAAP, Emergency Department, St. Christopher's Hospital for Children, PA

Eve Kimball, MD, FAAP, Southern California Permanente Medical Group

Paul Melinkovich, MD, FAAP, Kaiser Permanente

Jackie Nelson, MD, FAAP, Lander Regional H

Ellen Squire, MD, FAAP, HaysMed Pediatric Center

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 1999

Ad.3 Month and Year of most recent revision: 12, 2018

Ad.4 What is your frequency for review/update of this measure? Approximately every 3 years, sooner if the clinical guidelines have changed significantly.

Ad.5 When is the next scheduled review/update for this measure? 12, 2021

Ad.7 Disclaimers: 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.

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, “commercial use” refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code relating to the measures. Anyone desiring to use or reproduce the measures without modification for a noncommercial purpose may do so without obtaining approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA.