NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 0032

Corresponding Measures:

De.2. Measure Title: Cervical Cancer Screening

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

De.3. Brief Description of Measure: The percentage of women 21–64 years of age who were screened for cervical cancer using either of the following criteria:

-Women 21–64 years of age who had cervical cytology performed within the last 3 years.

-Women 30–64 years of age who had cervical high-risk human papillomavirus (hrHPV) testing performed within the last 5 years.

-Women 30–64 years of age who had cervical cytology/high-risk human papillomavirus (hrHPV) cotesting within the last 5 years.

1b.1. Developer Rationale: This measure assesses appropriate cervical cancer screening by seeking to ensure that women 21-64 years of age are screened for cervical cancer using the appropriate criteria for their age. Each year, approximately 12,000 women are diagnosed with cervical cancer in the U.S. (U.S. Cancer Statistics Working Group, 2015). Research suggests that cervical cancer is preventable with regular screening and follow-up and is curable if found and treated early. Adherence to this measure could lead to early treatment in affected women, which is associated with long survival and improved quality of life (CDC 2015).

Centers for Disease Control and Prevention (CDC). 2015. "Gynecologic Cancers: Cervical Cancer." http://www.cdc.gov/cancer/cervical/ (May 20, 2016).

U.S. Cancer Statistics Working Group. 2015. "United States Cancer Statistics: 1999–2012 Incidence and Mortality Web-based Report." Atlanta: U.S. Department of Health and Human Services. www.cdc.gov/uscs (May 20, 2016)

S.4. Numerator Statement: The number of women who were screened for cervical cancer.

S.6. Denominator Statement: Women 24-64 years of age as of the end of the measurement year.

S.8. Denominator Exclusions: This measure excludes women who had a hysterectomy with no residual cervix, cervical agenesis or acquired absence of cervix any time during their medical history through the end of the measurement year.

De.1. Measure Type: Process

S.17. Data Source: Claims, Electronic Health Data, Paper Medical Records

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Jan 17, 2017

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

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 U*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**

Summary of prior review in 2016

- The 2016 NQF Committee agreed the measure aligned with a USPSTF guideline that was based on comprehensive meta analyses.
- The developer established the focus of the measure (cervical cancer screening) as a secondary prevention measure to improve health outcomes via early detection of cervical cancer for treatment.
- The developer cited the 2012 USPSTF recommendations:

    - Women < 21 years – screening not recommended (Grade D Recommendation).
    - Women 21-29 years - screening with cytology alone every 3 years (Grade A Recommendation).
    - Women > 65 years – screening those who are not at high risk of cervical cancer and have had adequate prior screenings is not recommended (Grade D Recommendation).
    - Women after hysterectomy with removal of cervix – screening is not recommended (Grade D Recommendation).
- Recommendation with definition of grade:
    - Grade A: The USPSTF recommends the services. There is high certainty that the net benefit is substantial
    - 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.

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 previously provided evidence was a 2012 USPSTF guideline. For this submission, the developer provided an updated USPSTF guideline (2018).
- Previously, the guideline encompassed four recommended components. The 2018 guideline added: "Women 30-65 years: The USPSTF recommends screening with cytology alone every 3 years; or high-risk human papillomavirus (hrHPV) testing alone every 5 years; or cervical cytology/ hrHPV cotesting every 5 years (Grade A Recommendation)."

Exception to evidence

Not Applicable

Question for the Committee:

- The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency: high (Box 5) → High

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 2016 submission noted the average national performance in commercial plans was 75% and 60% in Medicaid plans.

- For the current submission, the developer provided the following commercial and Medicaid rates:
    - Commercial: Mean = 74.3%, 10th - 90th Percentile Range =66.1% - 82.1% (2019)
      Mean = 73.8%, 10th - 90th Percentile Range =65.6% - 81.8% (2018)
      Mean = 73.6%, 10th - 90th Percentile Range =65.5% - 81.4% (2017)
    - Medicaid: Mean = 59.3%, 10th - 90th Percentile Range =45.9% - 72.0% (2019)
      Mean = 59.4%, 10th - 90th Percentile Range =47.2% - 70.6% (2018)
      Mean = 58.0%, 10th - 90th Percentile Range =44.7% - 70.8% (2017)
- In 2018, HEDIS measures covered 116 million commercial health plan members and 54 million Medicaid enrollees.

Disparities

- The developer indicated that HEDIS data are stratified by type of insurance (e.g., Commercial, Medicaid, Medicare), but could be stratified by demographic variables.
- The developer provided additional literature addressing disparities and cervical cancer screening. Overall, less screening occurs in Hispanic and Asian populations.

Questions for the Committee:

- Is there a gap in care and/or disparities that warrant a national performance measure?
- Does this measure provide information to understand disparities in this area of healthcare?

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

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence

Comments:

- Good evidence, no concerns
- Process Measure
- The evidence, based on the 2018 USPSTF recommendation is very strong. It should be noted, however, that this recommendation is substantially different from the USPSTF 2012 recommendation, and the measure is substantially changed. In particular, for women aged 30-64, the USPSTF now gives two alternative screening methods: hrHPV testing alone or cervical cytology/hrHPV cotesting every 5 years. The brief measure information is not clear on how a new measure is formulated to take this into account. In particular "the number of women screened for cervical cancer" is not clear
- Rationale cites that screening and follow up support prevention of cervical cancer. Screening is well documented. Follow up is somewhat documented, though unsure whether, without results or follow up actions documented, the presence of screening truly leads to prevention.
- The measure is direct process measure of cervical cancer screening of women. The evidence is based on U.S. Preventive Services Task Force 2018 report on Screening for Cervical Cancer
- Very strong evidence supported by the most recent (2018) USPSTF guidance

1b. Performance Gap/Disparities

Comments:

- Disparities are apparent in the literature.
- Commercial and Medicaid results presented at mean, 10-90th percentile range. It is not stratified by demographic variables.
- There is clearly a performance gap using the previous measure, so the same is probably so for the new measure.
- Gap in care sufficiently demonstrated, particularly in Medicaid population
- Current performance data from 2018 was provided. It demonstrates a gap between the 25th to 75th percentile of 8-11 points in commercial and medicaid plans, respectively. Disparities are cited in the literature, not in direct evidence.
- There is demonstrable variability across payer types (commercial versus Medicaid). However, the developers do not explore known inequities (For example, certain Asian/Pacific Islander groups are less likely to be screened and there are also known inequities in screening for sexual minority women and women who live in rural areas). Stratification in the performance measure along those demographic lines (and incentives to close those gaps) is an important component of a QI-approach to screening

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

Question for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?

Question for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?

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

Measure Number: 0032

Measure Title: Cervical Cancer Screening

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.**
   - No concerns

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; Summary of Reliability Methodology
   - Beta-binomial model (ratio of signal to noise) was used to distinguish the performance of one accountable entity from another. The developer noted that the beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures, as is the case with most HEDIS measures.
   - The reliability testing was conducted at the measure score level; the ratio of signal to noise is an appropriate analysis for this measure at the health plan level.

7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3
   - Developer provided results of the reliability scores, as follows

     Distribution of Beta-Binomial Statistics, 2018

Plan Type	Overall Reliability
Commercial (402 plans, median 16,053 eligible patients per plan)	1.00
Medicaid (245 plans, median 21,447 eligible patients per plan)	0.99

   - The developer concluded that the reliability scores for commercial and Medicaid indicate that variation is caused by a real difference in performance (across accountable entities) rather than measurement error.

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.**

    - The values for the overall beta-binomial statistic across both product lines are 1.0 (maximum value possible) and 0.99, indicating the measure has high reliability.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    - Empirical testing of exclusions was not performed. The developer noted that the expert panel was engaged to inform the face validity of the measure exclusion, which aligns with evidence focused on cervical cancer screening for the general population of women ages 21-64.

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 was 8%, which represents an average of 4,010 additional women receiving cervical cancer screening in high-performing plans compared to low-performing plans. For Medicaid plans, the IQR was 11%, which represents an average of 3,700 additional women receiving cervical cancer screening in high-performing plans compared to low-performing plans.
    - 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.

    - Not applicable.

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 areaddressed 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 concerns.

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**

For cost/resource use measures ONLY:

17. **Are the specifications in alignment with the stated measure intent?**

☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**

20. **Method of establishing validity of the measure score:**

☐ **Face validity**

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

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

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

Submission document: **Testing attachment, section 2b2.2**

- The developer provided updated testing information. Empirical validity testing at the score level was conducted through construct validity to determine whether there was a correlation between:
  - *Cervical Cancer Screening* and the HEDIS *Breast Cancer Screening* measure, which assesses whether women 50-74 years had a mammogram to screen for breast cancer.
  - *Cervical Cancer Screening* and the HEDIS *Chlamydia Screening in Women* measure, which assesses whether women 16-20 years who were identified as sexually active had at least one test for chlamydia during the measurement year.
- The developer hypothesized that organizations that performed well on this measure would perform well on these two measures (i.e., positive correlation).
- The submission also described face validity testing for new measures.

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

Submission document: Testing attachment, section 2b2.3

- The developer reported the following results:
  - Pearson correlation coefficients ranged from 0.32-0.67 for the measure pairs for commercial health plans and Medicaid plans.
  - 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 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.

23. **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)

24. **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)

25. **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.)

26. **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

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

☐ **High**

☐ **Moderate**

☐ **Low**

☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

ADDITIONAL RECOMMENDATIONS

29. **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.**

- No concerns.

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

2a1. Reliability – Specifications

Comments:

- Good reliability
- Measure collected through claims data or manually extracted, consistently collected, no issue.
- How a single proportion based on the 2018 USPSTF recommendation (with differences in screening methods, age range, and screening interval) is created is not stated in the material that I have reviewed, so I cannot assess this.
- No concerns about implementation
- Data for reliability compared cervical cancer screening records using data abstraction and claims data, with clear data elements and calculation. No concerns.
- I have no concerns re: the reliability

2a2. Reliability – Testing

Comments:

- No
- No concerns, agree with high reliability rating.
- Beta-binomial testing was done, which is appropriate, but the resulting reliability statistics 1.00 for Commercial plans and 0.99 for Medicaid plans, look suspiciously high.
- No
- Reliability is high and estimated at between 0.99 to 1.0. No concerns about reliability.
- No

2b1. Validity – Testing

Comments:

- No
- Testing documentation submitted, missing data in HEDIS was addressed through an audit process.
- No concerns. Comparison with Breast Cancer and Chlamydia Screening is appropriate, and the values as expected.
- No
- Validity testing presented in the testing attachment raise no concerns.
- Validity testing seems to rest on the assumption that plans that perform well on one type of screening (breast cancer, in this example) will also perform well on other kinds of screening. I am not sure that is that case - the associations themselves are weak (.32) to moderate (.67). Would be interested to learn about the literature supporting the notion that from a woman's perspective "all screens are the same".

2b4-7. Threats to Validity

Comments:

- No
- Empirical testing of exclusions was not performed, however empirical validity testing of the measure score was included. Correlation measures to Breast Cancer screening and Chlamydia screening was assessed. Appropriate safeguards were included, no concerns.
- No concerns
- Possibly, particularly in Medicaid populations where members may come and go from coverage and health plan data may be incomplete
- No concerns based on potential threats to validity. Missing data are unlikely to constitute a threat.
- No

2b2-3. Other Threats to Validity
2b2. Exclusions
2b3. Risk Adjustment

Comments:

- Acceptable
- Risk Adjustment not applicable
- No concerns
- No concerns
- Exclusion of those with hysterectomy consistent with the evidence. Risk adjustment not performed.
- If we consider stratification a type of risk adjustment (as I do) then, no an appropriate strategy to unpack and highlight the inequities discussed above is not included in the measure.

Criterion 3. Feasibility

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

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

The developer reported:

- Data elements are generated and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition.
- Data elemenst are abstracted (e.g., chart abstraction for quality measure or registry) and coded (e.g., DRG, ICD-9 codes on claims) by someone other than person obtaining original information.
- Some data elements are available in defined fields in electronic sources.

Question for the Committee:

- Does the Committee have concerns about the feasiblity of this measure?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

- No concerns
- Data elements captured during provision of care, abstraction done by other personnel than healthcare provider. High rating for feasibility.
- No concerns
- No concerns
- Required data elements routinely generated and collected during care delivery. Use of claims data raises no concerns.
- No concerns

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

- The measure is used in several public reporting and payment programs, including:
  - California Align Measure Perform (Amp) Commercial HMO Program
  - California Align Measure Perform (Amp) Medi-Cal Managed Care Program
  - CMS Medicare Advantage Plan Rating System (“STARS”)
  - 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
- Qualified Health Plan (QHP) Quality Rating System (QRS)

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 reported the following:

- NCQA publishes HEDIS results annually and presents at various conferences and webinars.
- Technical assistance is provided on measures through the developer's Policy Clarification Support System.
- NCQA utilizes a consensus-based process to obrain broad input on the measure from several multi-stakeholder advisory panels, public comment posting, and questions submitted to the Policy Clarification Support System.
- In the past, feedback shared with the developer informed how the measure was revised to include new screening methods recommended by the U.S. Preventive Services Task Force and other major clinical guideline organizations.

***Question for the Committee**:*

- Has the measure been sufficiently 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 the following:

- Over the past three years, the measure has shown slight improvement (approximately 1% improvement over the past three years) across health plans. The greatest improvement in performance occurred for Medicaid plans (avg. 1.3% improvement in the average rate and 2% improvement for plans at the 90th percentile).
- An 11 point difference exists between Medicaid plans in the 25th and 75th percentiles in 2019, demonstrating room for improvement.

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 [unexpected findings]

Potential harms

- The developer did not report any unintended findings.

Questions for the Committee:

- 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

Comments:

- No concerns
- Publicly reported data and used in accountability programs (managed care, NCQA).
- No concerns
- No concerns; currently in regular use
- Used by NCQA in the HEDIS program and by others for public reporting. Feedback incorporated into the process.
- Yes, feedback is described.

4b1. Usability – Improvement

Comments:

- Good, no concerns
- Improvement noted in overall measure compliance, no unintended findings for failed compliance.
- No concerns
- No concerns
- The gaps in Medicaid plans indicate moderate potential for improvement through QI and other initiatives. Potential unintended concequences related to overuse in terms of higher than desirable frequency, but this is theoretical and most likely the plans have some procedures in place to reduce that.
- I don't perceive any unintended consequences

Criterion 5: Related and Competing Measures

Related or competing measures

- NQF 0579 : Annual cervical cancer screening or follow-up in high-risk women

Harmonization

- Specifications are not harmonized.
- The developer states that both measures focus on women who had cervical cancer screening during the year as the numerator, however, #0579 focuses on a denominator of high-risk patients and is used in a surveillance strategy.
- Exclusions are aligned across these measures.

Related and Competing Measures

5. Related and Competing

Comments:

- Not that I can see
- NQF 0579 Annual Cervical Cancer Screening or follow up in high risk women. These two measures are not harmonized. Major differnce is #579 is specifically targeted at high risk women. Exclusions are aligned across measures.
- NA
- One related measure that appears complementary; no concerns
- None known.
- None with the same denominator

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: June 30, 2020

- No NQF Members have submitted support/non-support choices as of this date.

Developer Submission

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

Evidence_Form_CCS.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.

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0032

Measure Title: Cervical Cancer Screening

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

Date of Submission: 4/16/2020

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

Outcome

☐ Outcome: Click here to name the health outcome

☐ Patient-reported outcome (PRO): Click here to name the PRO

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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Cervical Cancer Screening

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.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.

2020 Submission

Females at risk for cervical cancer >>> screening for cervical cancer >>> abnormal screening results >>> evaluation >>> early detection of abnormalities or cancer >>> treatment >>> improved length and/or quality of life.

2016 Submission

The measure focuses on a process (cervical cancer screening). The process, a secondary prevention measure, has been shown to improve outcomes by catching cervical cancer in its earlier, more treatable stages.

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.)

N/A

****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.

N/A

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

We have provided U.S. Preventive Services Task Force (USPSTF) recommendations published in 2018.

USPSTF Recommendation: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**2020 Submission** **U.S. Preventive Services Task Force. 2018. "Screening for Cervical Cancer U.S. Preventive Services Task Force Recommendation Statement." JAMA. 320(7):674-686. doi:10.1001/jama.2018.10897** **URL: https:// www.uspreventiveservicestaskforce.org/uspstf/ recommendation/ cervical-cancer-screening** **2016 Submission** **Moyer VA for the U.S. Preventive Services Task Force. 2012. Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 156(12):880-91.**
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.	• Women younger than 21 years: The USPSTF recommends against screening (Grade D Recommendation). • Women 21-29 years: The USPSTF recommends screening with cytology alone every 3 years (Grade A Recommendation). • Women 30-65 years: The USPSTF recommends screening with cytology alone every 3 years; or high-risk human papillomavirus (hrHPV) testing alone every 5 years; or cervical cytology/ hrHPV cotesting every 5 years (Grade A Recommendation). • Women older than 65 years: The USPSTF recommends against screening women who have had adequate prior screening and are not otherwise at high risk for cervical cancer (Grade D Recommendation). • Women after hysterectomy: The USPSTF recommends against screening in women who have had a hysterectomy with removal of the cervix and who do not have a history of a high-grade precancerous lesion or cervical cancer (Grade D Recommendation).
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	2020 Submission • The USPSTF concluded with high certainty that the benefits of screening every 3 years with cytology alone in women aged 21 to 29 years substantially outweigh the harms. • The USPSTF concluded with high certainty that the benefits of

	screening every 3 years with cytology alone, every 5 years with hrHPV testing alone, or in combination in women aged 30 to 65 years outweigh the harms. • The USPSTF concluded with moderate certainty that the benefits of screening in women older than 65 years who have had adequate prior screening and are not otherwise at high risk for cervical cancer do not outweigh the potential harms. • The USPSTF concluded with moderate certainty that the harms of screening in women younger than 21 years outweigh the benefits. • The USPSTF concluded with high certainty that the harms of screening in women who have had a hysterectomy with removal of the cervix for indications other than a high-grade precancerous lesion or cervical cancer outweigh the benefits.
Provide all other grades and definitions from the evidence grading system	N/A
Grade assigned to the **recommendation** with definition of the grade	• Grade A: The USPSTF recommends the services. There is high certainty that the net benefit is substantial • 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.
Provide all other grades and definitions from the recommendation grading system	• 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. • Grade C: The USPSTF recommends selectively offering or providing this service to individual patients based on professional judgment and patient preferences. There is at least moderate certainty that the net benefit is small. • I Statement: The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	2020 Submission The U.S. Preventive Services Task Force systematically reviewed the benefits and harms of screening for cervical cancer using hrHPV testing as the screening strategy (with or without cytology). They identified 8 large randomized controlled trials (4 of primary hrHPV testing and 4 of hrHPV contesting), 5 cohort studies and 1 individual

	participant data meta analysis. Trials were heterogeneous with regard to type of cytology (conventional vs. liquid-based cytology), type of hrHPV test (DNA PCR enzyme immunoassay vs. Hybrid Capture 2), screening interval (2 to 5 years), follow-up protocols for abnormal results, number of screening rounds (1 or 2), and consistency of screening protocols between rounds. Two fair-quality trials and one good-quality trial evaluated primary hrHPV screening (hrHPV testing alone) compared with cytology alone; two good- and two fair-quality trials compared hrHPV cotesting with cytology alone. 2016 Submission The measure is based on a USPSTF guideline that is based on a comprehensive meta-analysis (see USPSTF report for full number of studies).
Estimates of benefit and consistency across studies	2020 Submission The evidence was generally consistent in demonstrating that primary hrHPV testing increased detection of CIN3+ in the initial round of screening by as much as 2- to 3-fold. Evidence was mixed in cotesting trials. No trial showed a significant increase in CIN3+ detection in initial round of screening using cotesting. The decision model commissioned by the USPSTF reported that both hrHPV testing alone and cotesting would avert approximately 1 additional cancer case per 1,000 women screened compared with cytology alone (17.8 vs 16.5 cases, respectively), representing a very small improvement in life-years gained (64,193 vs 64,182 life-years, respectively). However, these 2 screening strategies would also subject women to more tests and procedures. Although no head-to-head trials compared screening with hrHPV testing alone vs cotesting, modeling suggests that both hrHPV testing alone and cotesting offer similar benefit over cytology in terms of cancer cases averted and are also similar in terms of the number of colposcopies required (1,630 vs 1,635, respectively). In summary, all 3 screening strategies offer substantial benefit in terms of reducing cancer incidence and mortality compared with no screening. 2016 Submission The USPSTF determined there was a positive net benefit and evidence was consistent.
What harms were identified?	2020 Submission Data to compare potential harms of different screening strategies were limited, and none of the included trials or observational studies

	reported on harms of the screening test or treatments. False-positive rates and referrals to colposcopy were in some trials 2- to 3-fold higher with hrHPV-based screening strategies relative to cytology alone in the first screening round, and evidence was lacking to determine whether these differences might persist over multiple screening rounds. The USPSTF concluded with high certainty that the benefits of screening every 3 years with cytology alone in women aged 21 to 29 years substantially outweigh the harms. The USPSTF concludes with high certainty that the benefits of screening every 3 years with cytology alone, every 5 years with hrHPV testing alone, or in combination in women aged 30 to 65 years outweigh the harms.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	N/A

_____

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 appropriate cervical cancer screening by seeking to ensure that women 21-64 years of age are screened for cervical cancer using the appropriate criteria for their age. Each year, approximately 12,000 women are diagnosed with cervical cancer in the U.S. (U.S. Cancer Statistics Working Group, 2015). Research suggests that cervical cancer is preventable with regular screening and follow-up and is curable if

found and treated early. Adherence to this measure could lead to early treatment in affected women, which is associated with long survival and improved quality of life (CDC 2015).

Centers for Disease Control and Prevention (CDC). 2015. “Gynecologic Cancers: Cervical Cancer.” http://www.cdc.gov/cancer/cervical/ (May 20, 2016).

U.S. Cancer Statistics Working Group. 2015. “United States Cancer Statistics: 1999–2012 Incidence and Mortality Web-based Report.” Atlanta: U.S. Department of Health and Human Services. www.cdc.gov/uscs (May 20, 2016)

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 are summarized at the health plan level by mean, standard deviation, and performance at the 10th, 25th, 50th, 75th and 90th percentile. We also calculated the interquartile range (IQR), which can be interpreted as the difference between the 25th?and 75th?percentile. Data are stratified by year and product line (i.e. commercial and Medicaid). The following data demonstrate room for improvement and variation in the rate of cervical cancer screening across health plans.

Commercial Rate

YEAR | MEAN | ST DEV | 10TH | 25TH | 50TH | 75TH | 90TH | Interquartile Range

2019 | 74.3% | 6.7% | 66.1% | 70.8% | 74.5% | 78.5% | 82.1% | 7.7

2018 | 73.8% | 7.3% | 65.6% | 70.4% | 74.4% | 78.0% | 81.8% | 7.6

2017 | 73.6% | 7.2% | 65.5% | 69.9% | 74.5% | 77.9% | 81.4% | 7.9

Medicaid Rate

YEAR | MEAN | ST DEV | 10TH | 25TH | 50TH | 75TH | 90TH | Interquartile Range

2019 | 59.3% | 11.6% | 45.9% | 55.2% | 60.5% | 66.2% | 72.0% | 10.9

2018 | 59.4% | 10.1% | 47.2% | 54.3% | 60.1% | 66.0% | 70.6% | 11.7

2017 | 58.0% | 11.4% | 44.7% | 51.9% | 58.4% | 65.7% | 70.8% | 13.8

These rates are extracted from HEDIS data collection and reflect the most recent years of measurement for this measure. For HEDIS 2019 (calendar year 2018), HEDIS measures covered 116 million commercial health plan members and 54 million Medicaid enrollees. Below is a description of the denominator for this measure. It includes the number of health plans included in HEDIS data collection and the mean eligible population for the measure across health plans.

Commercial

YEAR | N Plans | Mean Denominator Size

2019 | 402 | 20,108

2018 | 404 | 23,053

2017 | 418 | 14,237

Medicaid

YEAR | N Plans | Mean Denominator Size

2019 | 245 | 1,754

2018 | 262 | 936

2017 | 265 | 1,903

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, Medicare). 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. 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 measures 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.

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

In 2012, eight million U.S. women reported they had not been screened in the last 5 years (CDC 2014). Among women aged 21 to 65 years who responded to the 2015 National Health Interview Survey (NHIS), the national average reported rate for cytology alone was 81% and for cotesting was 32%. African American women were most likely to have had cervical cancer screening within 3 years; 85% of African American women reported having a cytology test and 35% reported having cotesting (Watson et al. 2017). The rates for white women were slightly above the national average rates (cytology alone: 83%; cotesting: 33%) (Watson et al. 2017). Both Hispanic women and Asian women had rates below the national average performance, with Asian women reporting the lowest screening rates (cytology alone: 74%; cotesting: 21%) (Watson et al. 2017). Multiple studies have found that barriers include Hispanic ethnicity, patient fear of finding cancer, and language could be barriers to screening (Akinlotan et al. 2017).

Despite gains among African American women (Watson et al. 2017; Beavis et al. 2017), a recent meta-analysis covering research from 2000 to 2012 found there is still a racial disparity in cervical cancer mortality. The mortality rate for African American women was 5.7 per 100,000, compared to 4.7 per 100,000 for white women (Beavis et al. 2017). Disparities in mortality still exist due to inadequate follow-up after screening, differences in treatment, and, in part, the higher-than-average rate of adenocarcinoma in African American

women (Galic et al. 2012; Wang et al. 2004). Adenocarcinoma is a rarer type of cervical cancer with malignant cells found in the inner part of the cervix (NCI 2018). Compared to squamous cell carcinoma, adenocarcinoma has a dramatically worse 5-year survival rate in stage II cervical cancer patients (Shimada et al. 2013), which may partially explain why African American women have a higher mortality rate.

Akinlotan M, Bolin JN, Helduser J, Ojinnaka C, Lichorad A, McClellan D. 2017. Cervical Cancer Screening Barriers and Risk Factor Knowledge Among Uninsured Women. J Community Health. 42(4): 770–778.

Beavis, A.L., Gravitt, P.E., Rositch, A.F. 2017. Hysterectomy-corrected cervical cancer mortality rates reveal a larger racial disparity in the United States. Cancer. 123(6):1044-50.

Centers for Disease Control and Prevention. 2014. "Cervical Cancer is Preventable." Last modified November 5, 2014 https://www.cdc.gov/vitalsigns/cervical-cancer/

Galic, V., Herzog, T.J., Lewin, S.N., et al. 2012. Prognostic significance of adenocarcinoma histology in women with cervical cancer. Gynecol Oncol. 125(2):287-91.

National Cancer Institute. 2018. "NCI Dictionary of Cancer Terms - Adenocarcinoma." (October 12, 2018) https://www.cancer.gov/publications/dictionaries/cancer-terms/def/adenocarcinoma

Shimada, M., Nishimura, R., Nogawa, T., Hatae, M., Takehara, K., Yamada, H. ... Kigawa, J. 2013. Comparison of the outcome between cervical adenocarcinoma and squamous cell carcinoma patients with adjuvant radiotherapy following radical surgery: SGSG/TGCU Intergroup Surveillance. Molecular and Clinical Oncology, 1, 780-784. (October 12, 2018) doi.org/10.3892/mco.2013.112

Wang, S.S., Sherman, M.E., Hildesheim, A., Lacey, J.V., Jr, Devesa, S. 2004. Cervical adenocarcinoma and squamous cell carcinoma incidence trends among white women and black women in the United States for 1976-2000. Cancer. 100(5):1035-44.

Watson, M., Benard, V., King, J., Crawford, A., Saraiya, M. 2017. National assessment of HPV and Pap tests: Changes in cervical cancer screening, National Health Interview Survey. Prev Med. 100:243-247. (October 12, 2018) doi: 10.1016/j.ypmed.2017.05.004.

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):*

Cancer, Cancer : Gynecologic

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

Primary Prevention, Screening

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

Populations at Risk

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

N/A

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 **:** 0032_CCS_Spring_2020_Value_Sets.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.

Yes

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

Since the last endorsement date, the U.S. Preventive Services Task Force released updated guidelines on cervical cancer screening and added a new screening method. Accordingly, NCQA updated the measure to align with the latest guidelines by adding high-risk HPV (hrHPV) testing alone every five years as an acceptable screening method for women ages 30-65. Additionally, as part of NCQA's annual measure maintenance, we routinely make coding and other minor specification updates to ensure the measure remains up-to-date with current practice and based on feedback received from measure users.

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

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

The number of women who were screened for cervical cancer.

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)*

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

ADMINISTRATIVE:

Number of women who were screened for cervical cancer through either of the following criteria:

-Women 24–64 years of age as of December 31 of the measurement year who had cervical cytology (Cervical Cytology Lab Test Value Set; Cervical Cytology Result or Finding Value Set) during the measurement year or the two years prior to the measurement year.

-Women 30–64 years of age as of December 31 of the measurement year who had cervical high-risk human papillomavirus (hrHPV) testing (High Risk HPV Lab Test Value Set, High Risk HPV Test Result or Finding Value Set) during the measurement year or the four years prior to the measurement year and who were 30 years or older on the date of the test.

NOTE: Evidence of hrHPV testing within the last 5 years also captures patients who had cotesting; therefore additional methods to identify cotesting are not necessary.

See attached value sets.

MEDICAL RECORD:

Number of women who were screened for cervical cancer through either of the following criteria:

-Women 24–64 years of age as of December 31 of the measurement year who had cervical cytology during the measurement year or the two years prior to the measurement year. Documentation in the medical record must include both of the following:

A note indicating the date when the cervical cytology was performed; and

The result or finding.

Count any cervical cancer screening method that includes collection and microscopic analysis of cervical cells. Do not count lab results that explicitly state the sample was inadequate or that “no cervical cells were present”; this is not considered appropriate screening.

Do not count biopsies because they are diagnostic and therapeutic only and are not valid for primary cervical cancer screening.

NOTE: Lab results that indicate the sample contained “no endocervical cells” may be used if a valid result was reported for the test.

-Women 30–64 years of age as of December 31 of the measurement year who had cervical high-risk human papillomavirus (hrHPV) testing during the measurement year or the four years prior to the measurement year and who were 30 years or older as of the date of testing. Documentation in the medical record must include both of the following:

A note indicating the date when the hrHPV test was performed. Generic documentation of “HPV test” can be counted as evidence of hrHPV test; and

The results or findings.

Do not count biopsies because they are diagnostic and therapeutic only and are not valid for primary cervical cancer screening.

NOTE: Evidence of hrHPV testing within the last 5 years also captures patients who had cotesting.

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

Women 24-64 years of age as of the end of the measurement year.

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

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

Use administrative data to identify all women 24-64 years of age as of the end of the measurement year.

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

This measure excludes women who had a hysterectomy with no residual cervix, cervical agenesis or acquired absence of cervix any time during their medical history through the end of 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.)*

ADMINISTRATIVE:

Exclude women who had evidence of hysterectomy with no residual cervix, cervical agenesis or acquired absence of cervix (Absence of Cervix Diagnosis Value Set, Hysterectomy with No Residual Cervix Value Set) any time during their medical history through the end of the measurement year.

See attached value sets.

MEDICAL RECORD:

Exclude women where there is documentation in the medical record of “complete,” “total” or “radical” abdominal or vaginal hysterectomy any time during their medical history through the end of the measurement year. The following also meet criteria:

-Documentation of a “vaginal pap smear” in conjunction with documentation of “hysterectomy.”

-Documentation of hysterectomy in combination with documentation that the patient no longer needs pap testing/cervical cancer screening. Documentation of hysterectomy alone does not meet the criteria because it is not sufficient evidence that the cervix was removed.

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

N/A

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

Step 1: Determine the eligible population: identify women 24-64 years of age as of the end of the measurement year.

Step 2: Exclude women who had evidence of hysterectomy with no residual cervix, cervical agenesis or acquired absence of cervix any time during their medical history through the end of the measurement year.

Step 3: Determine the numerator: identify the number of women who were screened for cervical cancer following the instructions in the numerator details listed in Section S.5.

Step 4: Divide the numerator from Step 3 by the denominator from Step 2 to determine the rate.

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.

This measure can be reported using Administrative and/or Medical Record data. For organizations that choose to report the measure using Medical Record data, a sample size of 411 is used. A sample size of 411 is used because it allows for the 95% confidence interval around the rate, meaning that a 5% difference in plan performance is statistically significant. NCQA provides a Random Number table that organizations use to assist with sample selection.

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, Electronic Health Data, Paper Medical Records

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data 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 and medical record documentation 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 Management Organizations and Preferred Provider Organizations 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. <u>COMPOSITE Performance Measure</u> - 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

Testing_Form_CCS-637226408173997699.docx

2.1 <u>For maintenance of endorsement</u>

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 <u>For maintenance of endorsement</u>

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 <u>For maintenance of endorsement</u>

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*)**:** 0032
Measure Title: Cervical Cancer Screening
Date of Submission: 1/13/2020

Type of Measure:

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

☐ Intermediate Clinical Outcome	☐ Cost/resource
☒ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.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: Click here to describe	☐ other: Click here to describe

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

2020 Submission N/A

2016 Submission N/A

1.3. What are the dates of the data used in testing? Click here to enter date range

2020 Submission 2018

2016 Submission 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: Click here to describe	☐ other: Click here to describe

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

2020 Submission

This measure assesses whether female health plan members ages 21-64 years were screened for cervical cancer using either of the following criteria: women 21-64 years who had cervical cytology performed within the last 3 years; women 30-64 years who had cervical high-risk human papillomavirus (hrHPV) testing performed within the last 5 years; or women 30-64 years who had cervical cytology/hrHPV cotesting within the last 5 years. Testing therefore was done at the health-plan level, which is appropriate for the level of reporting for this measure.

Measure score reliability and construct validity testing

Data used to assess reliability and validity were calculated from all Medicaid and commercial health plans submitting data to NCQA for this HEDIS measure. Data came from 245 Medicaid health plans and 402 commercial health plans that were geographically diverse and varied in size.

Systematic evaluation of face validity

The measure was assessed for face validity with three independent panels of experts.

- The Cervical Cancer Screening Measurement Advisory Panel included 6 experts in women's health, including representation by consumers, health plans, health care providers, and policy makers.
- The Technical Measurement Advisory Panel includes 12 members, including representation by health plan methodologists, clinicians, HEDIS auditors and state/federal users of measures.
- NCQA's Committee on Performance Measurement (CPM) 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 17 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

This question was not on the 2016 form.

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

2020 Submission

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

Table 1. Median eligible population for *Cervical Cancer Screening* by plan type, 2018

Plan Type	Number of Plans	Median number of eligible patients per plan
Commercial	402	16,053
Medicaid	245	21,447

2016 Submission

This question was not on the 2016 form.

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.

2020 Submission

No differences in the data used for reliability, construct validity and meaningful differences in performance testing. The systematic assessment of face validity was done with multi-stakeholder expert panels as described in Section 1.5 above.

2016 Submission

This question was not on the 2016 form.

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.

2020 Submission

Measure performance was assessed by commercial and Medicaid health plans, which serves as a proxy for socioeconomic status. We did not analyze additional social risk factors. Patient-reported data and patient community characteristics were not available in the testing data source.

2016 Submission

This question was not on the 2016 form.

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*)

2020 Submission

Reliability testing of performance measure score

We utilized the Beta-binomial model (Adams 2009) to assess how well one can confidently distinguish the performance of one accountable 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 in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS measures. 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 accountable entities).

In addition to the point estimate of signal-to-noise reliability, NCQA will also provide the standard error and 95% confidence interval (95% CI) by June 2, 2020. NCQA will include a summary of the methodology that was used to estimate the standard error and 95% CI.

Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

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*)

2020 Submission

Table 2. Estimated Signal-to-Noise Reliability for Commercial and Medicaid Plans, 2018

Plan Type	Overall Reliability
Commercial	0.99
Medicaid	0.99

* NCQA will provide the standard error and 95% CI for signal-to-noise reliability by June 2, 2020.

2016 Submission

The reliability for this measure remained high: for commercial plans, it was 1.00; for Medicaid plans, it was 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?*)

2020 Submission

The values for the overall beta-binomial statistic across both product lines are greater than 0.7, indicating the measure has very good reliability.

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)*

2020 Submission

Empiric Validity Testing of Performance Measure Score

We tested for construct validity by exploring the following:

- Is *Cervical Cancer Screening* correlated with the HEDIS *Breast Cancer Screening* measure, which assesses whether women 50-74 years had a mammogram to screen for breast cancer.
- Is *Cervical Cancer Screening* correlated with the HEDIS *Chlamydia Screening in Women* measure, which assesses whether women 16-20 years who were identified as sexually active had at least one test for chlamydia during the measurement year.

We hypothesized that organizations that perform well on *Cervical Cancer Screening* should perform well on the *Breast Cancer Screening* and *Chlamydia Screening in Women* measures. To test these correlations, we used a Pearson correlation test. This test estimates the strength of the linear association between two variables. The magnitude of correlation ranges from -1 to +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

Systematic Assessment of Face Validity of Performance Measure Score

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, 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. For the most recent updates to this measure in January 2019, the CPM voted to approve moving the proposed changes forward to public comment (15 CPM members approved, 0 members opposed and 0 abstained).

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. For the most recent updates to this measure in May 2019, the CPM voted to approve the measure for HEDIS health plan reporting (11 CPM members approved, 0 members opposed and 0 abstained).

2016 Submission

NCQA tested the measure for face validity using a panel of stakeholders with specific expertise in measurement This panel included representatives from key stakeholder groups, including specialists' in women's health, oncologists, family practitioners, health plans, state Medicaid agencies and researchers. 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.

The independent samples t-test is a measure of the probability that two population means are statistically significantly different.

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

2020 Submission

Table 3. Health-Plan Level Pearson Correlation Coefficients Among Cervical Cancer Screening and Other Measure Performance Scores in Commercial Plans, 2018

	Breast Cancer Screening	***Chlamydia Screening in Women***
Cervical Cancer Screening	0.67	0.50

**significant at p<0.001*

Table 4. Health-Plan Level Pearson Correlation Coefficients Among Cervical Cancer Screening and Other Measure Performance Scores in Medicaid Plans, 2018

	Breast Cancer Screening	***Chlamydia Screening in Women***
Cervical Cancer Screening	0.63	0.32

**significant at p<0.001*

2016 Submission

This measure was deemed valid by the expert panel.

The p value for commercial plans and Medicaid plans was <0.01.

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?*)

2020 Submission

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.

The correlation between performance on the *Cervical Cancer Screening* measure and the *Breast Cancer Screening* and *Chlamydia Screening in Women* measures was moderate. These results suggest that plans that perform well on one measure are moderately likely to perform well on the other measures.

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*)

2020 Submission

We did not perform empirical testing of the exclusion for hysterectomy with no residual cervix, cervical agenesis, or acquired absence of cervix for this submission. NCQA engaged expert panels to inform the face validity of this measure exclusion, which aligns with evidence focused on cervical cancer screening for the general population of women ages 21-64. This measure has been reviewed by NCQA's Cervical Cancer Screening Measurement Advisory Panel, Technical Measurement Advisory Panel, and the Committee on Performance Measurement. The measure also received public comment feedback upon initial development and during recent updates to the measure. Experts agreed that excluding individuals with no residual cervix, cervical agenesis or acquired absence of the cervix aligns is supported by clinical rationale.

2016 Submission

NCQA currently allows health plans for optional exclusion to their results. NCQA does not conduct the annual analysis applied to a sample. In measure development, field testing and any re-analysis for update, we investigate and validate the effect reliability 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*)

2020 Submission N/A

2016 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*)

2020 Submission N/A

2016 Submission N/A

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.

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

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

☐ **No risk adjustment or stratification**

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

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

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

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)*

To demonstrate meaningful differences in performance, NCQA calculated 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 calculated an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. 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 0.05, then the two plans' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans, one plan in the 25th percentile and another plan in the 75th percentile of performance. We used these two plans as examples of measures entities. However, the method can be used for comparison of any two measured entities.

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*)

2020 Submission

Table 5. Variation in Performance, 2018

Plan Type	No. of plans	Mean Eligible Population	Mean rate (%)	Min	10th	25th	50th	75th	90th	Max	IQR	p-value
Commercial	402	52,090	74.3	32.6	66.1	70.8	74.5	78.5	82.1	91.3	7.7	<0.0001
Medicaid	245	33,635	59.3	0.0	45.9	55.2	60.5	66.2	72.0	88.0	11.0	<0.0001

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).

Commercial Rate

YEAR| MEAN | ST DEV | 10TH | 25TH | 50TH | 75TH | 90TH | Interquartile Range

2014 | 75% | 6% | 68% | 71% | 76% | 79% | 82% | 7

2013 | 76% | 5% | 69% | 72% | 76% | 79% | 82% | 7
2012 | 75% | 5% | 69% | 72% | 75% | 77% | 80% | 6

Medicaid Rate
YEAR| MEAN | ST DEV | 10TH | 25TH | 50TH | 75TH | 90TH | Interquartile Range
2014 | 60% | 11% | 46% | 54% | 61% | 68% | 73% | 14
2013 | 62% | 12% | 46% | 55% | 64% | 71% | 76% | 17
2012 | 64% | 12% | 46% | 58% | 66% | 72% | 77% | 14

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 mean eligible population for the measure across health plans.

Commercial
YEAR | N Plans | Average Denominator Size
2014 | 405 | 21,417
2013 | 412 | 32,032
2012 | 416 | 44,559

Medicaid
YEAR | N Plans | Average Denominator Size
2014 | 220 | 1,423
2013 | 215 | 1,843
2012 | 195 | 1,073

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?*)

2020 Submission

For commercial plans, the IQR was 8%. This gap represents an average of 4,010 additional women receiving cervical cancer screening in high-performing plans compared to low-performing plans. For Medicaid plans, the IQR was 11%. This gap represents an average of 3,700 additional women receiving cervical cancer screening in high-performing plans compared to low-performing plans.

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.

This measure has only one set of specifications.

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 nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

2020 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*)

2020 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 nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

2020 Submission

All of the commercial and Medicaid health plans that reported 2018 HEDIS data for this measure reported valid rates as determined by NCQA-certified auditors through the process described above. 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 by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

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

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

Some data elements are in defined fields in electronic sources

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

To allow for widespread reporting across health plans, this measure is collected through multiple data sources (administrative data, electronic clinical data, paper records, and registry). We anticipate as electronic health records become more widespread the reliance on paper record review will decrease.

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 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 measure. This system informs both annual updates to the measures as well as routine re-evaluation of measures. These processes include updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence.

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)

	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/our-work/accountability/value-based-p4p California Align. Measure. Perform. (AMP) Medi-Cal Managed Care Program https://www.iha.org/our-work/accountability/medi-cal CMS Medicare Advantage Plan Rating System (STARS) https://www.medicare.gov/find-a-plan/questions/home.aspx California Align. Measure. Perform. (AMP) Commercial HMO Program https://www.iha.org/our-work/accountability/value-based-p4p California Align. Measure. Perform. (AMP) Medi-Cal Managed Care Program https://www.iha.org/our-work/accountability/medi-cal CMS Medicare Advantage Plan Rating System (STARS) https://www.medicare.gov/find-a-plan/questions/home.aspx 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.

CMS MEDICARE ADVANTAGE PLAN RATING SYSTEM ("STARS"): This measure is included in the composite Medicare Advantage Star Rating. CMS calculates a Star Rating (1-5) for all Medicare Advantage health plans based on 53 performance measures. Medicare beneficiaries can view the star rating and individual measure scores on the CMS Plan Compare website. The Star Rating is also used to calculate bonus payments to health plans with excellent performance. The Medicare Advantage Plan Rating program covers 11.5 million Medicare beneficiaries in 455 health plans across all 50 states.

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?*)

NA

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.*)

NA

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 Update and Best Practices Conference, 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. During this "reevaluation" process, we seek broad input on the measure, including input on performance and implementation experience. 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.

In general, health plans have not reported significant barriers to implementing this measure. Questions received through the Policy Clarification Support system have generally centered around minor clarification about the screening methods that satisfy the measure numerator. During a recent public comment session, a majority of comments from measured entities supported updates to the measure to align with the latest clinical recommendations.

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

This measure has been deemed a priority measure by NCQA and other entities such as the Centers for Medicare and Medicaid Services as illustrated by its use in the Medicare Advantage Health Plan Rating System and the 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 last major update, feedback obtained through the mechanisms described in 4a2.2.1 informed how we revised the measure to include new screening methods recommended by the U.S. Preventive Services Task Force and other major clinical guideline organizations.

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 health plans (see section 1b.2 for summary of data from health plans). The greatest improvement in performance has been seen for Medicaid plans (avg. 1.3% improvement in the

average rate and 2% improvement for plans at the 90th percentile). Additionally, in 2019 there was an 11 point difference between Medicaid plans in the 25th percentile and Medicaid plans in the 75th percentile, demonstrating additional room 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 consequences for this measure during testing or since implementation.

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

There were no identified unexpected findings 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)

0579 : Annual cervical cancer screening or follow-up in high-risk women

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?

No

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

The numerator for both measures focuses on women who had cervical cancer screening during the year, but #0579 focuses on a denominator of high-risk patients and is used in a surveillance strategy. The NCQA measure is intended to measure cervical cancer screening in the general population. Exclusions are aligned across these measures.

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.)
NA

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.

CERVICAL CANCER SCREENING MEASUREMENT ADVISORY PANEL

Andrea Gelzer, MD, AmeriHealth Caritas

Stephanie Glover, National Partnership for Women and Families

Debbie Saslow, PhD, American Cancer Society

George Sawaya, MD, University of California San Francisco

Laurie Spoll, Aetna

Kim Vesco, MD MPH, Kaiser Permanente Center for Health Research

COMMITTEE ON PERFORMANCE MEASUREMENT??

Andrew Baskin, MD, Aetna??

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 Hunter, (Co-Chair), MD, WPS Health Solutions?

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

Jeffrey Kelman,?MMSc, MD, Department of Health and Human Services

Nancy Lane, PhD, Independent Consultant

Bernadette Loftus, MD, The Permanente Medical Group

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

Amanda Parsons, MD, MBA, Montefiore Medical Center

Wayne Rawlins, MD, MBA, ConnectiCare

Rudy Saenz, MD, MMM, FACOG, Riverside Medical Clinic

Marcus?Thygeson, (Co-Chair), MD, MPH, Blind ON-Demand

JoAnn Volk, MA, Georgetown University

Lina Walker, PhD, AARP

NCQA TECHNICAL MEASUREMENT ADVISORY PANEL

Andy Amster, MSPH, Kaiser Permanente

Jennifer Brudnicki, MBA, Inovalon Inc.

Lindsay Cogan, PhD, MS, New York State Department of Health

Kathryn Coltin, MPH, Independent Consultant

Mike Farina, RPh, MBA, Capital District Physicians' Health Plan

Marissa Finn, MBA, CIGNA

Scott Fox, MS, MEd, FAMIA, The MITRE Corporation

Carlos Hernandez, CenCal Health

Harmon Jordan, ScD, Westat

Virginia Raney, LCSW, Center for Medicaid and CHIP Services

Lynne Rothney-Kozlak, MPH, Rothney-KozlakConsulting, LLC

Laurie Spoll, Aetna

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, Aqurate Health Data Management

Nelly Leon-Chisen, RHIA, American Hospital Association

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, WellCare Health Plans, Inc.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 1994

Ad.3 Month and Year of most recent revision: 12, 2019

Ad.4 What is your frequency for review/update of this measure? Approximately every three years

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 MEASURSE 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.

1100 13th Street NW, Third Floor
Washington, DC 20005
phone 202.955.3500
fax 202.955.3599
www.ncqa.org

Reliability Methodology for the Cervical Cancer Screening Measure

We utilized the methodology described by 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 in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures, such as the Cervical Cancer Screening 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 Cervical Cancer Screening measure, plans are the reporting entity.

The formula for signal-to-noise reliability is:

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}).

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, $\alpha > 0$, $\beta > 0$

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

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

n = number of eligible members (plan-specific denominator for the observed rate)

Using Adams' 2009 methodology, we estimated the reliability for each plan, then averaged these reliability estimates across all plans to produce a point estimate of signal-to-noise reliability. We are now labeling this point estimate "mean signal-to-noise reliability" to clarify its definition. The mean signal-to-noise reliability measures how well, on average, the measure can differentiate between plan performance on the Cervical Cancer Screening measure.

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

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 to provide additional information about the stability of reliability.

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 (σ2error) 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 to provide additional information about the distribution of plan-level signal-to-noise reliability estimates. The number of plans in each stratum and the number of eligible members per plan are displayed in the summary tables.

NCQA recently refined the methodology we use to calculate within-plan variance (σ2error). In prior submissions, σ2error was an estimate of the average error across all plans:

σ_error^2=(mean(p)*(1-mean(p)))/(mean(n))

p = observed rate for a plan

mean(p) = average rate for all plans

n = number of eligible members for that observed rate (plan denominator size for that measure)

mean(n) = average number of eligible members for all plans (average denominator size for that measure)

Our current methodology defines within-plan error as:

σ2error = $\hat{p}(1-\hat{p})/n$

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

n = number of eligible members (plan-specific denominator for the observed rate)

Our current methodology estimates σ2error for each plan rather than a single overall estimate of σ2error, allowing us to estimate the reliability for each plan and summarize the distribution of these estimates.

Signal-to-Noise Reliability Assessment for the Cervical Cancer Screening Measure

Table 1 compares the point estimates of mean signal-to-noise reliability using the current and former methodology. Although the current estimates are lower, they are above 0.90 for commercial and Medicaid plans, which is higher than the minimally accepted threshold of 0.70.

Table 1. Comparison of the Point Estimates of Mean Signal-to-Noise Reliability Using the Current and Former Methodology by Product Type, Calendar Year 2018 Data

	Point estimate: Mean Signal-To-Noise Reliability	
	Current Methodology	Former Methodology
Commercial	0.914	0.998
Medicaid	0.965	0.991

Table 2 provides the point estimate of mean signal-to-noise reliability, its standard error, and the 95% CI for the Cervical Cancer Screening measure for all commercial and Medicaid plans. For commercial plans, the estimate is 0.914, and the 95% CI is 0.907, 0.921. For Medicaid plans, the estimate is 0.965 and the 95% CI is 0.964, 0.967. The point estimate and lower bound of the 95% CI for both product types exceed the minimum threshold of 0.70, indicating that the measure has very good reliability.

Table 2 also provides the point estimate of mean signal-to-noise reliability, its standard error, and 95% CI for the Cancer Screening measure stratified by product type and denominator size (distribution of the number of eligible members per plan). Approximately 20% of commercial plans and 56% of Medicaid plans reported a denominator size of 411, indicating that these plans used hybrid reporting (a combination of administrative claims and medical chart review) for this measure. We took this information into account when creating strata for this analysis: denominator size < 411, 411 and >411. The number of plans in each stratum and the number of eligible members per plan are provided in Table 2. The results show that mean signal-to-noise reliability

estimates in each stratum for both commercial and Medicaid plans are well above 0.70, as is the lower bound of the 95% CI, indicating that the measure has very good reliability.

Table 2. Mean Signal-To-Noise Reliability for the Cervical Cancer Screening Measure by Denominator Size (Number of Eligible Members Per Plan) and for All Submissions Stratified by Plan Type, Calendar Year 2018 Data

Stratification	Number of Plans	Number of Eligible Members per Plan (min – max)	Mean Signal-To-Noise Reliability	SE	95% CI
All Commercial	402	36 – 665,937	0.914	0.003	0.907, 0.921
Group 1	186	36 – 410	0.821	0.004	0.813, 0.830
Group 2	79	411*	0.868	0.002	0.864, 0.871
Group 3	137	412 – 665,937	0.992	0.001	0.989, 0.995

Stratification	Number of Plans	Number of Eligible Members per Plan (min – max)	Mean Signal-To-Noise Reliability	SE	95% CI
All Medicaid	245	83 – 97,473	0.965	0.001	0.964, 0.967
Group 1	86	83 – 410	0.950	0.002	0.947, 0.954
Group 2	137	411*	0.933	0.001	0.933, 0.934
Group 3	22	412 – 97,473	0.998	0.001	0.997, 0.999

SE: Standard Error of the mean.

95% CI: 95% confidence interval.

Stratum of 411 eligible members per plan due to hybrid reporting (a combination of administrative claims and medical chart review) for 20% of commercial plans and 56% of Medicaid plans.

Table 3 provides the distribution of plan-level signal-to-noise reliability estimates. For all commercial plans, estimates range from 0.42 to 0.99; the 10th percentile (p10) is 0.86, which exceeds the 0.70 threshold for reliability. For Medicaid plans, estimates range from 0.89 to 0.99, indicating very good reliability.

Table 3 also includes the distribution of plan-level signal-to-noise reliability estimates stratified by denominator size. The 10th percentile of the distribution exceeds 0.80 for all strata for both commercial and Medicaid plans, indicating that the measure has very good reliability.

Table 3. Distribution of Plan Signal-To-Noise Reliability for the Cervical Cancer Screening Measure by Denominator Size (Number of Eligible Members Per Plan) and for All Submissions by Plan Type, Calendar Year 2018 Data

		Distribution of Plan Estimates of Signal-to-Noise Reliability						
Stratification	Number of Plans	Min	P10	P25	P50	P75	P90	Max
All Commercial	402	0.42	0.86	0.87	0.89	0.99	0.99	0.99
Group 1	186	0.34	0.81	0.82	0.83	0.84	0.85	0.92
Group 2	79	0.83	0.85	0.86	0.87	0.88	0.89	0.92
Group 3	137	0.90	0.98	0.99	0.99	0.99	0.99	0.99
All Medicaid	245	0.89	0.96	0.96	0.96	0.97	0.97	0.99

Group 1	86	0.86	0.95	0.95	0.95	0.95	0.96	0.97
Group 2	137	0.93	0.93	0.93	0.93	0.94	0.94	0.95
Group 3	22	0.99	0.99	0.99	0.99	0.99	0.99	0.99

*Groups 1-3 are defined in Table 2.