NATIONAL QUALITY FORUM
Driving measurable health improvements together

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through National Quality Forum's (NQF) Consensus Development Process (CDP). The information submitted by the measure developers/stewards is included after the *Brief Measure Information* and *Preliminary Analysis* 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 #: 1407
Corresponding Measures:
Measure Title: Immunizations for Adolescents
Measure Steward: National Committee for Quality Assurance (NCQA)
sp.02. Brief Description of Measure: Percentage of adolescents 13 years of age who had one dose of meningococcal conjugate vaccine, one tetanus, diphtheria toxoids and acellular pertussis (Tdap) vaccine, and have completed the human papillomavirus (HPV) vaccine series by their 13th birthday.
1b.01. Developer Rationale: Vaccines are critical tools for avoiding preventable illnesses in both the adolescent and general population. By encouraging vaccination of adolescent children, the measure protects these vulnerable individuals from avoidable morbidity and mortality while building important herd immunity and reducing medical costs.

---

sp.12. Numerator Statement: Adolescents who had at least one dose of meningococcal vaccine; at least one tetanus, diphtheria toxoids and acellular pertussis vaccine (Tdap); and the HPV vaccination series completed by their 13th birthday.
sp.14. Denominator Statement: Adolescents who turn 13 years of age during the measurement year.
sp.16. Denominator Exclusions: This measure excludes patients who have a contraindication for the vaccine and patients who use hospice services during the measurement year.

Measure Type: Process
sp.28. Data Source:
Electronic Health Records
Registry Data
Paper Medical Records
Claims
sp.07. Level of Analysis:
Health Plan

IF Endorsement Maintenance – Original Endorsement Date: 08/15/2011
Most Recent Endorsement Date: 6/7/2019

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement, endorsed measures are evaluated periodically to ensure that the measure 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 a change in evidence since the prior evaluation

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

The developer provides the following description for this measure:

- This is a maintenance process measure at the health plan level of analysis that measures the percentage of adolescents 13 years of age who had one dose of meningococcal conjugate vaccine, one tetanus, diphtheria toxoids and acellular pertussis (Tdap) vaccine, and have completed the human papillomavirus (HPV) vaccine series by their 13th birthday.
- The developer provides a logic model that depicts adolescents less than 13 years of age receiving meningococcal, HPV, and Tdap vaccinations, increasing resistance to bacterial diseases, and leading to improved health, length, and quality of life.

The developer provides the following evidence for this measure:

- SR 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 2018

- The measure is based on evidence from the Advisory Committee on Immunization Practices (ACIP) guidelines, which are endorsed by the Centers for Disease Control and Prevention (CDC).
- Since the Committee's last full review, Human Papilloma Virus (HPV) vaccination has been added to this measure and additional evidence was provided for the HPV vaccine. The ACIP graded the recommendation as evidence type three which means observational studies or randomized controlled trials with notable limitations, though the limitations were not provided.
- The developer also noted that no new additions were made to the body of evidence related to the meningococcal and Tdap/Td vaccines since the previous submission.
- The Standing Committee agreed that this measure met the evidence criterion.

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

Exception to evidence

- N/A

Questions for the Standing Committee:

- *The developer attests that the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Standing Committee agree that the evidence basis for the measure has not changed and that there is no need for repeated discussion and a vote on evidence?*
- *What is the relationship between this measure and patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

Guidance From the Evidence Algorithm

Process measure based on systematic review but not graded (Box 3) -> Empirical evidence submitted (Box 7) -> Summarized empirical evidence appears to include all studies (Box 8) -> Evidence indicates that benefits outweigh undesirable effects -> Moderate

Highest possible rating of Moderate.

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer reports the following data from the Healthcare Effectiveness Data and Information Set (HEDIS) reflecting the most recent years of measurement for this measure. Performance data is summarized at the health plan level and summarized by mean, interquartile range, and standard deviation. Data is stratified by year and product line.
    - Tdap
        - 2022
            - The commercial mean performance rate was 87.00 percent. The interquartile range (IQR) was 7.79 percent. The standard deviation was 7.13 percent.
            - The Medicaid mean performance rate was 83.39 percent. The IQR was 8.27 percent. The standard deviation was 8.27 percent.
        - 2021
            - The commercial mean performance rate was 86.11 percent. The IQR was 8.52 percent. The standard deviation was 8.67 percent.
            - The Medicaid mean performance rate was 85.76 percent. The interquartile range (IQR) was 7.54 percent. The standard deviation was 7.38 percent.
        - 2020

- The commercial mean performance rate was 87.15 percent. The IQR was 7.74 percent. The standard deviation was 7.69 percent.
- The Medicaid mean performance rate was 87.11 percent. The IQR was 6.32 percent. The standard deviation was 7.00 percent.

- HPV
  - 2022
    - The commercial mean performance rate was 33.72 percent. The IQR was 10.74 percent. The standard deviation was 9.32 percent.
    - The Medicaid mean performance rate was 37.74 percent. The IQR was 10.73 percent. The standard deviation was 9.16 percent.
  - 2021
    - The commercial mean performance rate was 31.62 percent. The IQR was 10.6 percent. The standard deviation was 9.20 percent.
    - The Medicaid mean performance rate was 39.86 percent. The IQR was 13.87 percent. The standard deviation was 9.74 percent.
  - 2020
    - The commercial mean performance rate was 29.89 percent. The IQR was 9.95 percent. The standard deviation was 9.42 percent.
    - The Medicaid mean performance rate was 39.98 percent. The IQR was 11.92 percent. The standard deviation was 10.17 percent.
- Meningococcal
  - 2022
    - The commercial mean performance rate was 82.43 percent. The IQR was 11.3 percent. The standard deviation was 8.65 percent.
    - The Medicaid mean performance rate was 79.38 percent. The IQR was 11.79 percent. The standard deviation was 9.49 percent.
  - 2021
    - The commercial mean performance rate was 81.31 percent. The IQR was 12.5 percent. The standard deviation was 9.47 percent.
    - The Medicaid mean performance rate was 81.91 percent. The IQR was 11.44 percent. The standard deviation was 9.33 percent.
  - 2020
    - The commercial mean performance rate was 81.42 percent. The IQR was 12.39 percent. The standard deviation was 9.64 percent.
    - The Medicaid mean performance rate was 82.97 percent. The IQR was 10.11 percent. The standard deviation was 9.33 percent.

Disparities

- The developer states that the measure can be stratified by demographic variables, such as race/ethnicity or socioeconomic status, to assess the presence of health care disparities, if the data are available to a plan.
- The developer cited evidence from the National Immunization Survey which they stated showed that national coverage with most routine childhood vaccines remained stable, however, disparities in

immunization coverage have been seen in uninsured, Black and Hispanic patients, and patients living below the federal poverty line compared to individuals who were privately insured, White, or living at or above the federal poverty line. The developer did not provide data to support this claim.

Questions for the Standing Committee:

- *Is there a gap in care that warrants a national performance measure?*

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

Criteria 2: Scientific Acceptability of Measure Properties

Complex measure evaluated by the Scientific Methods Panel (SMP)? ☐ **Yes** ☒ **No**

Evaluators: Staff

2a. Reliability: Specifications and Testing

For maintenance measures—no change in emphasis—specifications should be evaluated the same as with new measures.

2a1. Specifications require the measure, as specified, to produce consistent (i.e., reliable) and credible (i.e., valid) results about the quality of care when implemented.

For maintenance measures – less emphasis if no new testing data are provided.

2a2. Reliability testing demonstrates whether the measure data elements are repeatable and producing the same results a high proportion of the time when assessed in the same population during the same time period, and/or whether the measure score is precise enough to distinguish differences in performance across providers.

Specifications:

- Have the measure specifications changed since the last review? ☐ Yes ☒ No
- Measure specifications are clear and precise.

Reliability Testing:

- Did the developer conduct new reliability testing? ☒ Yes ☐ No
- Reliability testing conducted at the Accountable-Entity Level:
  - The developer conducted the same type of testing as they did in the previous submission with new data. The Standing Committee noted that no data element reliability testing was completed though the measure uses multiple data sources.
  - Reliability testing was conducted using signal to noise ratio (beta-binomial model) with 2018-2020 data. This was calculated for each of the 3 vaccines and stratified by commercial plans (N=391) and Medicaid plans (N=239). Reliability ranged from a low of 0.91 (HPV) to a high of 0.94 (Tdap/Meningococcal) among the commercial plans and from a low of 0.93 (Tdap) to a high of 0.95 (Meningococcal) among the Medicaid plans.
  - The average commercial plan reliability ranged from 0.92 (HPV) to 0.94 (Meningococcal/Tdap). The average Medicaid plan reliability ranged from 0.93 (Tdap) to 0.95 (Meningococcal).
  - The results suggest a high level of reliability for all 3 of the vaccine measures.

Questions for the Standing Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are the measure specifications adequate)?*
- *The developer attests that the specifications have not changed and that additional reliability testing was conducted but is directionally the same. Does the Standing Committee agree that the measure is still reliable and that there is no need for repeated discussion and a vote on reliability?*

Guidance From the Reliability Algorithm

Submitted specifications precise, unambiguous, and complete (Box 1) -> Empirical reliability testing conducted with the measure as specified (Box 2) -> Empirical reliability testing conducted at the accountable-entity level (Box 4) -> Reliability testing method described and appropriate (Box 5) -> High certainty or confidence that the accountable-entity levels are reliable (Box 6a) -> High

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

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

For maintenance measures – less emphasis if no new testing data are provided

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.

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

Validity Testing

- Did the developer conduct new validity testing? ☒ Yes ☐ No
- Validity testing conducted at the Accountable-Entity Level:
  - Validity was tested using construct validity by correlating vaccine rate for adolescents with vaccine rates for children under 2 years old.
    - The developer examined the correlation of Tdap with DTap and MMR, HPV with Rotavirus, and Meningococcal with VZV. These vaccines have similar dosing requirements.
    - The results were stratified by payer.
    - All correlations were positive. Commercial plan correlations ranged from 0.52 to 0.79 and Medicaid plan correlations ranged from 0.41 to 0.59. Statistical testing with a P value does not appear to be provided.

Exclusions

- This measure excludes patients who have a contraindication for the vaccine and patients who use hospice services during the measurement year.
- The developer did not conduct statistical analyses to determine the impact of the exclusions on the measure rates, noting low rates of reported plan exclusions.
- The developer did provide information on prevalence of the exclusions by payer:
  - For commercial plans, 91 of 391 plans reported any exclusions and among those reporting exclusions 0.69% of their population was excluded on average.
  - Among Medicaid plans, 16 of 239 plans reported any exclusions and among those reporting exclusions 0.02% of the population were excluded.

Risk Adjustment

- The measure is not risk adjusted.
- The developer concluded there is no conceptual reason to risk-adjust a measure assessing vaccination rates.

Meaningful Differences

- The developer calculated an interquartile range for each of the 3 vaccine rates and all 3 combined and then conducted an independent sample t-test between two randomly selected plans in each group (below 25th percentile and above 75th percentile).
- The IQR ranged from 9 to 13 percentage points for commercial plans and from 8 to 14 percentage points for Medicaid plans.
- The developer states all measures for both commercial and Medicaid plans were statistically different from zero using a p-value threshold of 0.05.

Missing Data

- NCQA has an audit process to check to ensure the HEDIS measures are correctly identified and reported.
- If a data source is found to have missing data and the issue cannot be rectified, the auditor will assign a "materially biased" designation to the measure for the reporting plan and the rate will not be reported.

Comparability

- The measure only uses one set of specifications for this measure.

Questions for the Standing Committee regarding validity:

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

Guidance From the Validity Algorithm

All potential threats to validity that are relevant to the measure empirically assessed (Box 1) à Empirical validity testing conducted using the measure as specified and with appropriate statistical testing (Box 2) -> Empirical validity testing conducted at the accountable entity level for each level of analysis specified in the measure (Box 5) -> Validity testing method described and appropriate (Box 6) -> Moderate certainty or confidence that the accountable entity levels are a valid indicator of quality (Box 7b) -> Moderate.

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

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 reports that the data elements needed to compute the measure are generated or collected by and used by healthcare personnel during the provision of care and are coded by someone other than the person obtaining original information.
- The developer states that some data elements are in defined fields in electronic sources.
- The developer reports that to allow for widespread reporting across health plans and health care practices, if applicable, this measure is collected through multiple data sources including administrative data, electronic clinical data, paper records, and registry data.
- The developer noted that the measure is not currently developed as an eCQM.
- The developer states that the measure's current use and history of use demonstrate its feasibility for reporting entities. Data collection from medical records, administrative sources, and registry data is robust and well-supported.

The developer encourages broad public use and dissemination of the measure for noncommercial uses. However, commercial use of the measure requires the prior written consent of NCQA.

Questions for the Standing Committee:

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

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

Criterion 4: Use and Usability

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

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

Current uses of the measure

Publicly reported?	☒	Yes	☐	No		
Current use in an accountability program?	☒	Yes	☐	No	☐	UNCLEAR
Planned use in an accountability program?	☐	Yes	☒	No	☐	NA

Accountability program details

- The developer reports that the measure is publicly reported and currently in use in various accountability programs.
    - The developer states that the NCQA Health Plan Rating program uses the measure to calculate health plan ratings, which are reported in Consumer Reports and on the NCQA website. These

ratings are based on performance on HEDIS measures, and in 2011, 646 Medicare health plans, 576 commercial health plans, and 278 Medicaid health plans across 50 states were included.

- The developer reports that the measure is publicly reported nationally and by geographic regions in the NCQA Annual State of Health Care Quality 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.
- The developer reports that the measure is used in the Centers for Medicaid and Medicare Services (CMS) Medicaid Child Core Set. The data collected from the measure will help CMS better understand the quality of health care that children enrolled in Medicaid and the Children's Health Insurance Program (CHIP) receive nationally.
- The developer reports that the measure is used in the CMS Health Insurance Marketplaces Quality Rating System.
- The developer reports that the measure is used in scoring for accreditation of Medicare Advantage Health Plans in the NCQA Health Plan Accreditation program. In 2019, 336 commercial health plans covering 87 million people and 77 Medicaid health plans covering 9.1 million people were accredited. Health plans are scored based on performance compared to benchmarks.
- The developer reports that the measure is used in the Quality Compass Program, which provides a tool used for selecting a health plan, conducting competitor analysis, examining quality improvement, and benchmarking plan performance.

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

Feedback on the measure provided by those being measured or others

- The developer reports that NCQA publicly reports rates across all plans and creates benchmarks to help health plans understand how they perform relative to others.
- The developer states that HEDIS results are published annually in the Quality Compass tool. Additionally, the developer presents data at various conferences and webinars and provides technical assistance through its Policy Clarification Support System.
- The developer reports that NCQA measures are evaluated regularly using a consensus-based process to consider input from multiple stakeholders, including but not limited to entities being measured. Additionally, the developer states that several methods are used to solicit input, including vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and a review of questions submitted to the Policy Clarification Support System.

***Questions for the Standing Committee**:*

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

Preliminary rating for Use: ☒ Pass ☐ No Pass

4b. Usability (4b1. Improvement; 4b2. Benefits of measure)

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

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

Improvement results

- The developer reports that the number of accountable entities has increased for this measure.
  - Previous submission data showed from 2012-2014 an average of commercial plans reporting was 346 and for Medicaid it was 175 plans.
  - Data submitted in 2019-2021 shows an average of 391 commercial plans and 240 Medicaid plans reporting.
- The developer reports that performance rates for this measure stayed high with some fluctuation, and that rate fluctuation may be a result of the COVID-19 pandemic.

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

Unexpected findings (positive or negative) during implementation

- The developer states that there were no identified unintended findings for this measure during testing or since implementation.

Potential harms

- The developer states that there were no identified unexpected benefits for this measure during testing or since implementation.

Additional Feedback:

- The measure was reviewed by the Measure Applications Partnership in 2013. The Workgroup recommended support for the measure to be added to Physician Compare/Value-Based Payment Modifier (VBPM) and for it to remain included in Physician Quality Reporting System (PQRS) as it was NQF-endorsed.

***Questions for the Standing Committee**:*

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

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

Criterion 5: Related and Competing Measures

Related/Competing Measures

- The developer does not identify any NQF-endorsed or non-NQF endorsed related or competing measures.
- NQF staff identified the following related measures:
  - NQF #0038 Childhood Immunization Status (CIS)
  - NQF #3620 Adult Immunization Status
  - NQF #3484 Prenatal Immunization Status

Criteria 1: Importance to Measure and Report

1a. Evidence

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

---

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

[Response Begins]

No

[Response Ends]

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

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01. Provide a logic model.

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

[Response Begins]

Previous Submission:

Adolescents less than 13 years of age >> meningococcal, HPV and Tdap vaccinations are performed >> increased resistance to bacterial diseases >> improved health, length and quality of life

[Response Ends]

1a.02. Select the type of source for the systematic review of the body of evidence that supports the performance measure.

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

[Response Begins]

Clinical Practice Guideline recommendation (with evidence review)

[Response Ends]

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, add additional tables by clicking "Add" after the final question in the group.

Evidence - Systematic Reviews Table (Repeatable)

Group 1 - Evidence - Systematic Reviews Table

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

[Response Begins]

Previous Submission:

Centers for Disease Control and Prevention. 2018. Prevention of Pertussis, Tetanus, and Diphtheria with Vaccines in the United States: Recommendations of the Advisory Committee on Immunization Practices (ACIP). https://www.cdc.gov/mmwr/volumes/67/rr/pdfs/rr6702a1-H.pdf

[Response Ends]

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

[Response Begins]

Previous Submission:

"Persons aged 11–18 years should receive a single dose of Tdap, preferably at a preventive care visit at ages 11–12 years."

[Response Ends]

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

[Response Begins]

Previous Submission:

ACIP did not provide a grade for the evidence underlying this recommendation. ACIP conducts a thorough review of peer-reviewed evidence on vaccine safety and effectiveness, discusses recommendations with professional organizations, and holds regular meetings for experts to vote on proposed recommendations.

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

Previous Submission:

ACIP did not provide a grade for this recommendation. ACIP conducts a thorough review of peer-reviewed evidence on vaccine safety and effectiveness, discusses recommendations with professional organizations, and holds regular meetings for experts to vote on proposed recommendations.

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

Previous Submission: The ACIP Pertussis Vaccines Work Group reviewed the "epidemiology of pertussis, tetanus, and diphtheria in the United States; use of Tdap vaccine among persons aged ≥65 years, children aged 7–10 years, health care personnel, and women during pregnancy; minimum interval between the last tetanus toxoid–containing vaccine and receipt of Tdap; effectiveness of Tdap vaccine; and vaccine safety." The Work Group reviewed literature from 2004 to 2017, covering topics such as vaccine effectiveness, Tdap revaccination, and post-licensure safety.

[Response Ends]

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

[Response Begins]

Previous Submission: The Work Group reviewed vaccine safety—compared to the morbidity and mortality risks associated with pertussis, tetanus and diphtheria—and concluded that "All persons are recommended to receive routine pertussis, tetanus, and diphtheria vaccination. Vaccine type, product, number of doses and booster dose recommendations are based on age and pregnancy status." This includes routine vaccination of adolescents at ages 11–12 years with a single dose of Tdap.

[Response Ends]

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

[Response Begins]

Previous Submission: The Work Group examined the risk of adverse events for each vaccine component, including hypersensitivity reactions, and concluded that the Tdap vaccine is safe for routine administration. ACIP identified two contraindications to the Tdap vaccine: severe allergic reactions or encephalopathy associated with administration of a prior dose of a DTP, DTap, or Tdap vaccine.

[Response Ends]

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

[Response Begins]

Previous Submission: There have been no studies published since the guideline that would significantly affect the findings.

[Response Ends]

Group 2 - Evidence - Systematic Reviews Table

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

[Response Begins]

Previous Submission:

Centers for Disease Control and Prevention. 2013. Prevention and Control of Meningococcal Disease: Recommendations of the Advisory Committee on Immunization Practices (ACIP). https://www.cdc.gov/mmwr/pdf/rr/rr6202.pdf

[Response Ends]

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

[Response Begins]

Previous Submission:

“ACIP recommends routine vaccination with a quadrivalent meningococcal conjugate vaccine (MenACWY) for adolescents aged 11 or 12 years, with a booster dose at age 16 years.”

[Response Ends]

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

[Response Begins]

Previous Submission:

ACIP did not provide a grade for the evidence underlying this recommendation. ACIP conducts a thorough review of peer-reviewed evidence on vaccine safety and effectiveness, discusses recommendations with professional organizations, and holds regular meetings for experts to vote on proposed recommendations.

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

Previous Submission:

ACIP did not provide a grade for this recommendation. ACIP conducts a thorough review of peer-reviewed evidence on vaccine safety and effectiveness, discusses recommendations with professional organizations, and holds regular meetings for experts to vote on proposed recommendations.

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

Previous Submission:

ACIP's Meningococcal Vaccines Work Group reviewed data on the "safety, efficacy, and immunogenicity of meningococcal vaccines." The Work Group uses "published, peer-reviewed studies" as well as unpublished data on immunogenicity and postlicensure observational data.

[Response Ends]

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

[Response Begins]

Previous Submission:

The Work Group reviewed vaccine safety—compared to the burden of meningococcal disease—and concluded that "all persons aged 11 through 18 years" should be vaccinated with MenACWY.

[Response Ends]

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

[Response Begins]

Previous Submission: The Work Group identified potential adverse events associated with vaccination, including injection site-swelling, fever, headache, and nausea.

[Response Ends]

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

[Response Begins]

There have been no studies published since the guideline that would significantly affect the findings.

[Response Ends]

Group 3 - Evidence - Systematic Reviews Table

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

[Response Begins]

Previous Submission:

Centers for Disease Control and Prevention. 2014. Human Papillomavirus Vaccination: Recommendations of the Advisory Committee on Immunization Practices (ACIP). https://www.cdc.gov/mmwr/pdf/rr/rr6305.pdf

Centers for Disease Control and Prevention. 2016. Use of a 2-Dose Schedule for Human Papillomavirus Vaccination — Updated Recommendations of the Advisory Committee on Immunization Practices. https://www.cdc.gov/mmwr/volumes/65/wr/pdfs/mm6549a5.pdf

[Response Ends]

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

[Response Begins]

"ACIP recommends routine HPV vaccination at age 11 or 12 years. Vaccination can be given starting at age 9 years." (CDC 2016)

[Response Ends]

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

[Response Begins]

3-dose schedule: No grade (the practice of grading ACIP recommendations was not established until after the release of this recommendation) (CDC 2014)

2-dose schedule: Evidence type 3: Observational studies or randomized controlled trials with notable limitation. (CDC 2016)

[Response Ends]

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

[Response Begins]

Evidence Type 1: Randomized controlled trials or overwhelming evidence from observational studies.

Evidence Type 2: Randomized controlled trials with important limitations, or exceptionally strong evidence from observational studies.

Evidence Type 4: Clinical experience and observations, observational studies with important limitations or randomized controlled trials with several major limitations.

[Response Ends]

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

[Response Begins]

"ACIP recommends a 2-dose schedule for HPV vaccination of girls and boys who initiate the vaccination series at ages 9 through 14 years (Category A recommendation)" (CDC 2016)

Category A: Recommendation that applies to all persons in an age or risk-based group.

[Response Ends]

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

[Response Begins]

Category B: Recommendation for individual clinical decision making.

[Response Ends]

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

[Response Begins]

The HPV Vaccine Work Group reviewed data on "safety, immunogenicity, and efficacy" of the HPV vaccine, as well as "data on epidemiology and natural history of HPV, sexual behavior, vaccine acceptability, and cost-effectiveness of HPV vaccination" (CDC 2014). The Work Group also reviewed the "immunogenicity, efficacy, and postlicensure effectiveness of a 2-dose schedule" (CDC 2016).

[Response Ends]

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

[Response Begins]

ACIP concluded that "HPV vaccines are highly effective and safe, and a powerful prevention tool for reducing HPV infections and HPV-associated cancers" (CDC 2016).

[Response Ends]

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

[Response Begins]

The Work Group identified potential adverse events associated with vaccination, including syncope, nausea, injection site pain, and fever.

[Response Ends]

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

[Response Begins]

There have been no studies published since the guideline that would significantly affect the findings.

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

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

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

[Response Begins]

Vaccines are critical tools for avoiding preventable illnesses in both the adolescent and general population. By encouraging vaccination of adolescent children, the measure protects these vulnerable individuals from avoidable morbidity and mortality while building important herd immunity and reducing medical costs.

[Response Ends]

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

Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

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

HPV performance rates for commercial and Medicaid plans, 2020 -2022.

Measurement Year	Plan Type	Total Number of Plans (N)	Mean	Standard Deviation	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
2022	Commercial	390	33.74%	9.32%	23.14%	27.46%	32.76%	38.20%	45.08%
2021	Commercial	386	31.62%	9.20%	20.92%	25.90%	30.70%	36.50%	43.80%
2020	Commercial	383	29.89%	9.42%	19.49%	23.87%	28.82%	33.82%	42.09%
2022	Medicaid	247	37.74%	9.16%	27.25%	31.85%	36.50%	42.58%	51.20%
2021	Medicaid	249	39.86%	9.74%	28.71%	32.36%	38.44%	46.23%	52.55%
2020	Medicaid	242	39.98%	10.17%	27.98%	33.58%	38.81%	45.50%	53.28%

Meningococcal performance rates for commercial and Medicaid plans, 2020 -2022

Measurement Year	Plan Type	Total Number of Plans (N)	Mean	Standard Deviation	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
2022	Commercial	390	82.43%	8.65%	71.93%	77.86%	83.70%	89.16%	92.09%
2021	Commercial	386	81.31%	9.47%	68.40%	76.06%	83.44%	88.56%	91.73%
2020	Commercial	383	81.42%	9.64%	67.73%	76.17%	83.14%	88.56%	92.11%
2022	Medicaid	247	79.38%	9.49%	66.67%	74.83%	80.78%	86.62%	89.29%
2021	Medicaid	249	81.91%	9.33%	69.10%	77.37%	84.18%	88.81%	90.91%
2020	Medicaid	242	82.27%	9.93%	69.35%	78.94%	84.91%	89.05%	91.48%

Tdap performance rates for commercial and Medicaid plans, 2020 -2022

Measurement Year	Plan Type	Total Number of Plans (N)	Mean	Standard Deviation	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
2022	Commercial	390	87.00%	7.13%	77.44%	84.18%	88.52%	91.97%	93.93%
2021	Commercial	386	86.11%	8.67%	73.72%	83.45%	88.55%	91.97%	94.16%
2020	Commercial	383	87.15%	7.69%	76.53%	84.91%	89.05%	92.65%	94.39%
2022	Medicaid	247	83.39%	8.27%	73.24%	80.29%	85.20%	88.56%	91.48%
2021	Medicaid	249	85.76%	7.38%	78.10%	82.97%	87.46%	90.51%	92.46%
2020	Medicaid	242	87.11%	7.00%	80.05%	85.16%	88.65%	91.48%	93.43%

Total numbers of plans and average denominator size for commercial and Medicaid plans, 2019 - 2021

Measurement Year	Plan Type	Total Number of Plans	Average Denominator Size
2021	Commercial	386	1,200
2020	Commercial	383	983
2019	Commercial	377	886
2021	Medicaid	249	516
2020	Medicaid	242	545
2019	Medicaid	222	627

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.

[Response Begins]

N/A

[Response Ends]

1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

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. The HEDIS Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing, and using race/ethnicity and language data to assess health care disparities.

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.

[Response Begins]

Variations in immunization coverage exist among some populations. Data from the National Immunization Survey showed that national coverage with most routine childhood vaccines remained stable. However, disparities in coverage have been seen in uninsured patients, Black and Hispanic patients, and patients living below the federal poverty line compared to individuals who were privately insured, White, or living at or above the poverty line (Hill et al., 2021).

(Hill, Holly A., et al. *Vaccination Coverage by Age 24 Months Among Children Born in 2017 and 2018 – National Immunization Survey-Child, United States, 2018-2020*. No. 41, 2021, p. 6.)

[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

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

---

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

[Response Begins]

No

[Response Ends]

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

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

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

[Response Begins]

We have not made any important changes to the measure specifications since the last measure review.

[Response Ends]

sp.01. Provide the measure title.

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

[Response Begins]

Immunizations for Adolescents

[Response Ends]

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

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

[Response Begins]

Percentage of adolescents 13 years of age who had one dose of meningococcal conjugate vaccine, one tetanus, diphtheria toxoids and acellular pertussis (Tdap) vaccine, and have completed the human papillomavirus (HPV) vaccine series by their 13th birthday.

[Response Ends]

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

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

Please do not select:

- *Surgery: General*

[Response Begins]

Infectious Diseases (ID)

[Response Ends]

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

[Response Begins]

Immunization

Primary Prevention

[Response Ends]

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

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

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

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Children (Age < 18)

[Response Ends]

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

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

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

Please do not select:

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

[Response Begins]

Health Plan

[Response Ends]

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

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

[Response Begins]

Outpatient Services

[Response Ends]

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

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

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

Available in attached Excel or csv file

[Response Ends]

Attachment: 1407_1407_1407 IMA Fall 2022 Value Sets-508.xlsx

sp.13. State the numerator.

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

DO NOT include the rationale for the measure.

[Response Begins]

Adolescents who had at least one dose of meningococcal vaccine; at least one tetanus, diphtheria toxoids and acellular pertussis vaccine (Tdap); and the HPV vaccination series completed by their 13th birthday.

[Response Ends]

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

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

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

[Response Begins]

ADMINISTRATIVE: For meningococcal, Tdap and HPV, count only evidence of the antigen or combination vaccine.

Meningococcal: At least one meningococcal vaccine (Meningococcal Vaccine Administered Value Set), with a date of service on or between the member's 11th and 13th birthdays.

Tdap: At least one tetanus, diphtheria toxoids and acellular pertussis (Tdap) vaccine (Tdap Vaccine Administered Value Set) with a date of service on or between the member's 10th and 13th birthdays.

HPV: At least two HPV vaccines (HPV Vaccine Administered Value Set), with different dates of service on or between the member's 9th and 13th birthdays. There must be at least 146 days between the first and second dose of the HPV vaccine. OR At least three HPV vaccines (HPV Vaccine Administered Value Set), with different dates of service on or between the member's 9th and 13th birthdays.

All Vaccines (Meningococcal, Tdap, HPV): Adolescents who are numerator compliant for all three indicators (meningococcal, Tdap, HPV).

MEDICAL RECORD: For meningococcal, Tdap and HPV, count only evidence of the antigen or combination vaccine.

For immunization information obtained from the medical record, count members where there is evidence that the antigen was rendered from either of the following: a, a note indicating the name of the specific antigen and the date of the immunization, or b, a certificate of immunization prepared by an authorized health care provider or agency, including the specific dates and types of immunizations administered.

For the two-dose HPV vaccination series, there must be at least 146 days between the first and second dose of the HPV vaccine.

For meningococcal vaccination, do not count serogroup B (MenB) vaccines. Immunizations documented under a generic header of "meningococcal" and generic documentation that the "meningococcal vaccine" was administered meet criteria.

Immunizations documented using a generic header of "Tdap/Td" can be counted as evidence of Tdap.

[Response Ends]

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

Adolescents who turn 13 years of age during the measurement year.

[Response Ends]

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

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

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

[Response Begins]

Step 1: Identify adolescents who turned 13 years of age during the measurement year

Step 2: Remove those who are not enrolled 12 months prior to the 13th birthday

Step 3: Remove members with a gap in enrollment of 46 days or longer during the 12 months prior to the 13th birthday. Exclude members not enrolled on the member's 13th birthday. Members in hospice or using hospice services anytime during the measurement year. Members with a documented contraindication for the vaccine may be excluded if the contraindication was documented before the member's 13th birthday.

Step 4: Remove all required exclusions listed in sp.18

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

This measure excludes patients who have a contraindication for the vaccine and patients who use hospice services during the measurement year.

[Response Ends]

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

All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

Any of the following on or before the member's thirteenth birthday meet exclusion criteria:

- Children in hospice or using hospice services
- Severe combined immunodeficiency (Severe Combined Immunodeficiency Value Set)
- Immunodeficiency (Disorders of the Immune System Value Set)
- HIV (HIV Value Set; HIV Type 2 Value Set)
- Lymphoreticular cancer, multiple myeloma or leukemia (Malignant Neoplasm of Lymphatic Tissue Value Set).
- Intussusception (Intussusception Value Set).

[Response Ends]

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

Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

No

[Response Ends]

sp.21. Select the risk adjustment type.

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

[Response Begins]

No risk adjustment or risk stratification

[Response Ends]

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

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

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

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

[Response Begins]

Better quality = Higher score

[Response Ends]

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

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

[Response Begins]

Step 1. Determine the eligible population: identify adolescents 13 years of age by the end of the measurement year.

Step 2. Exclude patients who had an anaphylactic reaction to the vaccines or its components.

Step 3: Determine the numerator: identify the number of patients who have received the meningococcal vaccine, Tdap vaccine, and HPV vaccine series.

Step 4. Calculate a rate for each individual vaccination as well as combinations of vaccinations (All vaccine rate: Tdap, meningococcal, and HPV).

[Response Ends]

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

Examples of samples used for testing:

- *Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.*
- *The sample should represent the variety of entities whose performance will be measured. The 2010 Measure Testing Task Force recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.*
- *The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.*
- *When possible, units of measurement and patients within units should be randomly selected.*

[Response Begins]

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 can use to assist with sample selection.

[Response Ends]

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

[Response Begins]

Claims

Electronic Health Records

Paper Medical Records

Registry Data

[Response Ends]

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

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

[Response Begins]

This measure is based on administrative claims and medical record documentation collected in the process 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.

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

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

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

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

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

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

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

[Response Begins]

No

[Response Ends]

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

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

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

[Response Begins]

No additional risk adjustment analysis included

[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

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

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

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

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

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

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

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

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

[Response Begins]

Claims

Electronic Health Records

Paper Medical Records

Registry Data

[Response Ends]

2a.02. If an existing dataset was used, identify the specific dataset.

The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

2022 Submission: 01-01-2018 – 12-31-2020

[Response Ends]

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

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

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

Please do not select:

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

[Response Begins]

Health Plan

[Response Ends]

2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).

Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.

[Response Begins]

Testing was done at the health plan level, which is appropriate for this measure. Data used to assess reliability were calculated from all Medicaid and commercial health plans submitting data to NCQA for this HEDIS measure. Data came from 239 Medicaid health plans and 391 commercial health plans that were geographically diverse and varied in size.

[Response Ends]

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

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

[Response Begins]

Data in the analysis came from 239 Medicaid health plans and 391 commercial health plans with a diverse national patient population. 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. The HEDIS Race/Ethnicity Diversity of Membership and the Language Diversity of Membership measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing, and using race/ethnicity and language data to assess health care disparities.

[Response Ends]

2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.

[Response Begins]

The same dataset was used for all testing.

[Response Ends]

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

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

[Response Begins]

This measure is specified to be reported separately by Medicaid and commercial plan types, which serves as a proxy for income and other socioeconomic factors.

[Response Ends]

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

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

Choose one or both levels.

[Response Begins]

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

[Response Ends]

2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.

Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.

[Response Begins]

Methodology described by John Adams (Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009) was used to calculate signal-to-noise reliability. Reliability was estimated by using the beta-binomial model. This model assesses how well one can confidently distinguish the performance of one reporting entity to another. For HEDIS measures, the health plan is the reporting entity.

The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures. Reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error, whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance across reporting entities. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another.

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}).

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

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

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

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

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

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

Along with the point estimate of mean signal-to-noise reliability, we are also providing the distribution of the plan-level (and provider-level) signal-to-noise reliability estimates. Each reporting unit's reliability estimate is a ratio of signal to noise, as described above [$\sigma^2_{plan\text{-}to\text{-}plan}$ / ($\sigma^2_{plan\text{-}to\text{-}plan}$ + σ^2_{error})]. Variability between reporting units ($\sigma^2_{plan\text{-}to\text{-}plan}$) is the same for each unit, while the specific reporting unit error (σ^2_{error}) varies. Reliability for each reporting unit is an ordinal measure of how well one can determine where that entity lies in the distribution across reporting units, with higher estimates indicating better reliability.

[Response Ends]

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

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

[Response Begins]

Table 3a shows the reliability for the overall measure as shown by the beta-binomial model. Table 2b shows the variability of individual plan reliability.

Overall Reliability

Measure Rate	Commercial	Medicaid
HPV	0.91	0.94
Meningococcal	0.94	0.95
Tdap	0.94	0.93

Individual plan reliability for the measure rates of HPV, Meningococcal, and Tdap by commercial and Medicaid reliability.(Average, 10th percentile, Median, 90th percentile)

Measure Rate	Commercial Avg	Commercial 10th	Commercial 50th	Commercial 90th	Commercial Avg	Commercial 10th	Commercial 50th	Commercial 90th
HPV	0.92	0.86	0.9	0.98	0.94	0.93	0.94	0.95
Meningococcal	0.94	0.89	0.96	0.99	0.95	0.93	0.96	0.98
Tdap	0.94	0.88	0.96	0.98	0.93	0.90	0.94	0.96

[Response Ends]

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

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

[Response Begins]

In general, a score of 0.7 or higher suggests the measure has adequate reliability. The results suggest the measure has very good reliability.

[Response Ends]

2b. Validity

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

[Response Begins]

Empirical validity testing

[Response Ends]

2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.

Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.

[Response Begins]

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

as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

CIS and IMA were compared to each other along the following indicator sets:

CIS Indicator	IMA Indicator	Rationale
DTaP	Tdap	Both assess the same type of vaccine
MMR	Tdap	Similar dosing requirements
Rotavirus	HPV	Similar dosing requirements
VZV (Varicella)	Meningococcal	Similar dosing requirements

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]

Commercial

CIS Indicator	IMA Indicator	Correlation
DTaP	Tdap	0.79
MMR	Tdap	0.67
Rotavirus	HPV	0.52
VZV (Varicella)	Meningococcal	0.59

Medicaid

CIS Indicator	IMA Indicator	Correlation
DTaP	Tdap	0.59
MMR	Tdap	0.55
Rotavirus	HPV	0.41
VZV (Varicella)	Meningococcal	0.54

[Response Ends]

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

[Response Begins]

The pairs of indicators are all positively associated with each other, across both product lines. Correlations were moderate to high across the pairs. The results indicate that as health plans improve rates for one measure, rates for the other also improve, which is reasonable given the similarities between the measures.

[Response Ends]

2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.

Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure. To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected reporting units from each group (below 25th and above 75th percentiles). The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each reporting unit. The test statistic is then compared against a normal distribution. If the p-value of the test statistic is less than .05, then the two reporting units' performance are significantly different from each other.

[Response Ends]

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

Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.

[Response Begins]

Commercial and Medicaid by measure rates

Measure Rate	Commercial 25th	Commercial 75th	Commercial p-value	Medicaid 25th	Medicaid 75th	Medicaid p-value
Combination [SB1] 2	0.243274854	0.348387097	0.0000	0.309002433	0.435523114	p < 0.001
HPV	0.258957655	0.364963504	0.0101	0.323600973	0.462287105	p < 0.001
Meningococcal	0.76056338	0.885644769	0.0000	0.773722628	0.888077859	p < 0.001
Tdap	0.834549878	0.919708029	0.0000	0.829683698	0.905109489	p < 0.001

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

[Response Ends]

2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.

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

[Response Begins]

Statistically significant and meaningful variation in performance across plans exists for this measure.

[Response Ends]

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

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

[Response Begins]

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

[Response Ends]

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

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

[Response Begins]

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. 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.

[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.

[Response Begins]

This measure goes through the NCQA audit process each year to identify potential errors or bias in results. Only performances rates that have been reviewed and determined not to be "materially biased" are reported and used.

[Response Ends]

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eCQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

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

[Response Begins]

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

[Response Ends]

2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.

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

[Response Begins]

[Response Ends]

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

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

Yes, the measure uses exclusions.

[Response Ends]

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

Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?

[Response Begins]

Due to low rates of reported plan exclusions, exclusions were not tested by individual exclusion criteria (i.e., data for those excluded by hospice, data for those excluded by vaccine components). Exclusions had a minimal overall effect on performance rates for the measure.

[Response Ends]

2b.17. Provide the statistical results from testing exclusions.

Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.

[Response Begins]

Commercial and Medicaid by plan type

Plan Type	2020
Commercial	91 plans out of 391 plans reporting excluded patients for clinical reasons. Among those plans, only 0.69% of their eligible population, on average, was excluded.
Medicaid	16 plans out of 239 plans reporting excluded patients for clinical reasons. Among those plans, 0.02% of their eligible population, on average, was excluded.

The very small number of exclusions reported by plans did not have a discernible impact on measure performance scores.

[Response Ends]

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

In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.

[Response Begins]

Given the very small number of exclusions across all reporting plans, exclusions for allergy or intolerance to the vaccine have a minimal effect on the overall performance rates for the measure. However, the exclusions are still necessary because they remove patients for clinical reasons.

[Response Ends]

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

[Response Begins]

No risk adjustment or stratification

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

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

[Response Begins]

[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

Criterion 3. Feasibility

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

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

[Response Begins]

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

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

[Response Ends]

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

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

Some data elements are in defined fields in electronic sources

[Response Ends]

3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.

[Response Begins]

To allow for widespread reporting across health plans (and health care practices, if applicable), 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.

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

Immunizations for Adolescents is not currently developed as an eCQM.

[Response Ends]

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

[Response Begins]

This measure's current use and history of use demonstrate its feasibility for reporting entities. Data collection from medical records, administrative sources, and registry data is robust and well-supported.

[Response Ends]

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

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

Attach the fee schedule here, if applicable.

[Response Begins]

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.

[Response Ends]

Criterion 4: Use and Usability

4a. Use

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

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

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

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

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

[Response Begins]

Public Reporting

[Public Reporting Please Explain]

Program Name: NCQA Health Plan Rating

- **URL**: https://reportcards.ncqa.org/health-plans
- This measure is used to calculate health plan ratings, which are reported in Consumer Reports and on the NCQA website. These rankings are based on performance on HEDIS measures among other factors. In 2021, a total of 643 Medicare health plans, 576 commercial health plans and 278 Medicaid health plans across 50 states were included in the rankings.

Program Name: NCQA Annual State of Health Care Quality

• **URL:** https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-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.

Program Name: CMS Medicaid Child Core Set

- **URL**: https://www.medicaid.gov/medicaid/quality-of-care/downloads/2022-child-core-set.pdf
- These are a core set of health quality measures for children enrolled in Medicaid/Children's Health Insurance Program (CHIP) to be reported at the state level. The data collected from these measures will help CMS to better understand the quality of health care that children enrolled in Medicaid/CHIP receive nationally.

Program Name: CMS Health Insurance Marketplaces - Quality Rating System

- **URL**: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/ACA-MQI/Quality-Rating-System/About-the-QRS
- The Affordable Care Act requires that qualified health plans participating in the Health Insurance Marketplaces submit quality rating information, including clinical measures. Data will be publicly reported.

Payment Program

Regulatory and Accreditation Programs

[Regulatory and Accreditation Programs Please Explain]

Program Name: NCQA Health Plan Accreditation

- **URL**: https://www.ncqa.org/programs/health-plans/health-plan-accreditation-hpa/

This measure is used in scoring for accreditation of Medicare Advantage 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.

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

[Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Please Explain]

Program Name: Quality Compass

- **URL:** https://www.ncqa.org/programs/data-and-information-technology/data-purchase-and-licensing/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.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Measure Currently in Use

[Response Ends]

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

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

[Response Begins]

N/A

[Response Ends]

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

A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

[Response Begins]

N/A - measure is in use

[Response Ends]

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

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

[Response Begins]

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.

[Response Ends]

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

[Response Begins]

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

[Response Ends]

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

[Response Begins]

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

[Response Ends]

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

[Response Begins]

Questions received through the Policy Clarification Support system have generally centered around clarification on the interval between HPV vaccine doses and allowable documentation as proof of vaccination.

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.

[Response Ends]

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

[Response Begins]

This measure has been deemed a priority measure by NCQA and other entities, as illustrated by its use in programs such as the CMS Medicaid Child Core Set.

[Response Ends]

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

[Response Begins]

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 updated recommendations for adolescent vaccines from the CDC's Advisory Committee on Immunization Practices.

[Response Ends]

4b. Usability

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

[Response Begins]

This measure was first introduced in HEDIS in 2009, and rates for HPV vaccination were added in 2016.

The number of accountable entities has increased for this measure. Previous submission data showed from 2012-2014 an average of commercial plans reporting was 346 and for Medicaid was 175 plans. Data listed above in the Importance to Measure and Report: Gap in Care/Disparities section shows an average of 391 commercial plans and 240 Medicaid plans reporting for 2019-2021. Performance rates for this measure generally stayed high with some fluctuation. Rate fluctuation may be a result of to COVID-19 pandemic. A study by Onimoe et al. identified the impact on COVID-19 on Well Child Care and Vaccination. Using medical record review, it was found that 43.5% of patients within 2020 were not up to date on their childhood vaccinations.

(Onimoe, Grace, et al. "Effect of COVID-19 Pandemic on Well Child Care and Vaccination." *Frontiers in Pediatrics*, vol, 10, Apr. 2022, p. 873482. PubMed Central, https://doi.org/10.3389/fped.2022.873482)

[Response Ends]

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

[Response Begins]

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

[Response Ends]

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

[Response Begins]

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

[Response Ends]

Criterion 5: Related and Competing Measures

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

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

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

(Can search and select measures.)

[Response Begins]

[Response Ends]

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

(Can search and select measures.)

[Response Begins]

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

[Response Begins]

No

[Response Ends]

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

[Response Begins]

N/A

[Response Ends]

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

Provide analyses when possible.

[Response Begins]

N/A

[Response Ends]