NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3484

Corresponding Measures:

De.2. Measure Title: Prenatal Immunization Status

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

De.3. Brief Description of Measure: Percentage of deliveries in the measurement period in which women received influenza and tetanus, diphtheria toxoids and acellular pertussis (Tdap) vaccinations.

1b.1. Developer Rationale: See question 1c.3.

S.4. Numerator Statement: Deliveries in which women received influenza and tetanus, diphtheria toxoids and acellular pertussis (Tdap) vaccinations.

S.6. Denominator Statement: Deliveries that occurred during the measurement period.

S.8. Denominator Exclusions: Deliveries that occurred at less than 37 weeks gestation.

Deliveries in which women were in hospice during the measurement period.

De.1. Measure Type: Composite

S.17. Data Source: Claims, Electronic Health Data, Electronic Health Records, Enrollment Data, Management Data, Other, Registry Data

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

IF this measure is included in a composite, NQF Composite#/title: The percentage of deliveries in the measurement year in which women had received influenza and tetanus, diphtheria and acellular pertussis (Tdap) vaccinations.

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not applicable.

Criteria 1: Importance to Measure and Report

1a. Evidence

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

The developer provides the following evidence for this measure:

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

Evidence Summary:

- Developer cited guidelines for each of the prenatal vaccines which are referenced in the measure description and the Advisory Committee on Immunization Practices (ACIP) recommendations. A brief description of the body of evidence for each vaccine is below
- **Tetanus, Diphtheria Toxoids and Acellular Pertussis (Tdap) Vaccine Recommendation.** The ACIP Pertussis Vaccines Work Group reviewed available data and evidence covering topics such as tetanus, diphtheria and pertussis disease epidemiology in the United States, decision analyses, cost-effectiveness, programmatic considerations, vaccine immunogenicity, vaccine safety, and postlicensure Tdap vaccine effectiveness. In total, they reviewed 110 studies consisting of randomized control trials and other types of studies for Tdap and Td vaccination, and approximately 77 studies on prenatal Tdap vaccination (including 16 studies on prenatal Tdap vaccine effectiveness and 5 studies on timing of Tdap vaccination during pregnancy).
- **Influenza Vaccine Recommendation.** ACIP Influenza Work Group reviewed data on immunogenicity, efficiency, effectiveness and safety of influenza vaccines. In total, the Work Group reviewed approximately 285 studies on the immunogenicity, efficacy and effectiveness of IIV, RIV and LAIV and 120 studies on influenza vaccine safety. This review includes approximately 12 studies specifically assessing immunogenicity, efficacy and effectiveness of influenza vaccine for pregnant women; and 46 studies assessing influenza vaccine safety for pregnant women. These studies consist of randomized control trials, case control studies and observational studies, among others.

Questions for the Committee:

- *Does the Committee agree the evidence basis for the measure?*

Guidance from the Evidence Algorithm

Measure a health outcome (Box 1) No → Assess performance of intermediate outcome, process, or structure(Box 3) Yes → Empirical evidence without SR or QQC (Box 4) yes → Grade for evidence (Box 5a) Yes → High

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

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

Maintenance measures – increased emphasis on gap and variation

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

The developer cited data extracted from HEDIS data collection reflecting the most most recent year of measurement (2018) for this measure. Performance data is summarized at the health plan level and summarized by mean performance and performance at 10th, 25th, 50th, 75th, and 90th percentile. The interquartile range (IQR) was also extracted, which can be interpreted as the difference between the 25th and 75th percentile. Data is stratified by product line (i.e. commercial, Medicaid and Medicare).

The following data demonstrate the variation in the rate of prenatal immunization across health plans.

For the indicator assessing receipt of influenza vaccination among pregnant women, there was a 12 point difference between plans in the 25th percentile and plans in the 75th percentile for commercial plans and 11 points for Medicaid plans. For the indicator assessing receipt of Tdap vaccination among pregnant women, there was a 17 point difference between plans in the 25th percentile and plans in the 75th percentile for commercial plans and 16 points for Medicaid plans. These gaps in performance underscore the opportunity for improvement.

Prenatal Immunization Status: Both Vaccines

Commercial, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

33.1 | 18.4 | 26.6 | 33.6 | 39.0 | 44.5 | 12.4

Medicaid, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

16.7 | 8.1 | 12.2 | 17.0 | 19.6 | 25.3 | 7.4

Prenatal Immunization Status: Influenza

Commercial, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

40.5 | 27.3 | 33.3 | 40.7 | 45.5 | 52.4 | 12.2

Medicaid, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

23.8 | 13.1 | 17.2 | 23.5 | 28.0 | 32.2 | 10.8

Prenatal Immunization Status: Tdap

Commercial, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

62.7 | 44.5 | 55.2 | 65.4 | 72.2 | 77.3 | 17.0

Medicaid, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

40.4 | 27.2 | 33.3 | 40.6 | 48.8 | 56.3 | 15.5

The HEDIS performance data reflect the most recent year of measurement for this measure. Below is a description of the number of health plans that reported this measure and the median denominator for the measure (stratified by commercial and Medicaid).

Commercial, 2018

N Plans | Median Denominator Size

68 | 1,374

Medicaid, 2018

N Plans | Median Denominator Size

19 | 3,800

Disparities

The developer cited that HEDIS data are stratified by type of insurance and while disparities data was not specified in the measure, it can 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.

Questions for the Committee:

- If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?

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

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

1a. Evidence to Support Measure Focus

- ACIP recommendation Clinical Practice Guidelines.
- Evidence appears to be directly related, based on ACIP guidelines for vaccines during pregnancy.
- Strong evidence has been provided.
- The evidence is appropriate.
- The recommendations for each specific vaccine during pregnancy are well supported by evidence-based recommendations from ACIP. Rating: High
- High rating for evidence to support the measure focus

1b. Performance Gap

- Results were stratified by Insurer type Commercial vs. Medicaid and documented a performance gap. Although the measure could be stratified by race/ethnicity this anlaysis was not done. Disparities were summarized from the literature.
- Current data provided. Appears to be less than optimal performance in each block analyzed. Data by carrier type provided, but no demographic detail. Variation between carriers is apparent.
- Evidence is provided for gaps both within and between payor strata. However, given the strength of the literature re: maternal/child health inequities (including vaccination- https://www.sciencedirect.com/science/article/pii/S0264410X19302087) I am surprised no effort to investigate or at least specific a conceptual rationale was attempted. Payor type is a very blunt proxy for socioeconomic status
- I agree with the moderate rating.
- The submission focuses on variation among health plans as the primary evidence of performance gaps. While this is important, it should also be noted how low the current rates are across the board for vaccines that are universally recommended. The submission does not address disparities, but does present data documenting differences between Commercial, Medicaid, and Medicare plans, which can be a proxy for disparities. I strongly suspect that there are major disparities by race, ethnicity, education, income, geography and other socio-demographic characteristics. Thus, I

do believe that there are major performance gaps to address, the submission does not do a good job of presenting them. Rating: Moderate
- Performance gap data appears to demonstrate gaps in care, while offering an opportunity to measure disparities in care with this specific measure and overall maternity care.

1c. Composite Performance Measure

- All or none weighting for this composite measure Influenza and TDAP are combined to produce the composite measure
- This composite makes sense. I see no reason to weight.
- No concerns noted.
- 1c. Composite Performance Measure - Quality Construct (if applicable): Are the following stated and logical: overall quality construct, component performance measures, and their relationships; rationale and distinctive and additive value; and aggregation and weighting rules? The rationale for the composite measure is straightforward: each of the ACIP-recommended prenatal vaccines are included. This rationale is not, however, addressed in the submission. It should be noted that the definition of the composite is unusual: it is neither an all-or-nothing measure nor an average. Rather, it is defined for each age group as: Denominator: total number of recommended vaccines in the population monitored Numerator: total number of vaccines received in the same population For example, if there were a group of 100 individuals in an age-group in which 2 vaccines were recommended, and 20 and 40 individuals had received vaccines A and B respectively, the measure would be 60/200 = 0.30. Note that although this is a proportion, it does not have a binomial distribution since any individual's receiving vaccine A and B are probably not independent events. Rating: Moderate
- N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

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

Validity

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

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

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

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

Evaluators:

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Reliability

RELIABILITY: SPECIFICATIONS

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

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

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

2. **Briefly summarize any concerns about the measure specifications.**

Panel member 1: There inconsistency between how the measure has been specified specifically in sections (S.5) and (S.14). Unless the measure require patients to have continuous enrollment from July 1 of prior year through the measurement year, how would it capture the immunization history of the plan members.

"Numerator 1: Deliveries where members received an influenza vaccine on or between July 1 of the year prior to the measurement period and the delivery date; or deliveries where members had an influenza virus vaccine adverse reaction any time during or before the Measurement Period". (S.5)

"Determine the eligible population. Identify all deliveries during the measurement period (January 1 – December 31) in which the patient was continuously enrolled from 28 days prior to delivery through the delivery date." (S.14)

Panel member 2: No response

Panel member 3: SMO: No concerns about specifications

Panel member 4: None

Panel member 5: No Concerns

Panel member 6: None.

RELIABILITY: TESTING

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

3. **Reliability testing level ☒ Measure score ☐ Data element ☐ Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure ☒ Yes ☒ No**
   Panel member 4: The measure developer indicated integrated delivery system as a level of analysis; I did not see any data that demonstrated testing with integrated delivery systems.

5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

☐ **Yes** ☐ **No**

6. **Assess the method(s) used for reliability testing**

Panel member 1: Adam's Beta Binomial model which captures signal-to-noise to assess reliability.

Panel member 2: Beta-binomial model-ratio of signal to noise was appropriate
Panel member 3: SMO: The developers used a beta- binomial model to estimate signal-to-noise reliability.

Panel member 4: Used beta-binominal for signal-to-noise, which is appropriate given that the measure is a pass/fail measure.

Panel member 5: The developer used a signal-to-noise analysis, which was appropriate for score level analysis of reliability.

Panel member 6: Beta-binomial model

Submission document: Testing attachment, section 2a2.2

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

Panel member 1: High level (>0.99.) of overall reliability as measured by Beta Binomial coefficient.

Panel member 2: Demonstrated strong statistical reliability

Panel member 3: SMO: The results indicate near perfect reliability

Panel member 4: All of the beta-binominal statistic across all product lines is greater than 0.7, indicating the measure has good reliability.

Panel member 5: Arrived at nearly perfect reliability

Panel member 6: Among the 68 commercial and 19 Medicaid plans, reliability of the measure was excellent, with median reliability estimates exceeding 0.98 for both commercial and Medicaid plans for the measure component and the composite measure.

Submission document: Testing attachment, section 2a2.3

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

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

Submission document: Testing attachment, section 2a2.2

☒ **Yes**

☐ **No**

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

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

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

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

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

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

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    Panel member 1: As indicated in #2 above, I am still not clear about how this quality measure will be operationalized unless it speficies the continuous enrollment requirement in the health plan.

Panel member 2: No concerns, submitors analysis supported high reliability
Panel member 3: SMO: My high rating of reliability was based on the near perfect signal-to-noise ratio estimate and no major concerns about the specifications.

Panel member 4: The measure developer indicated integrated delivery system as a level of analysis, but it appears as if they did not provide testing of the measure with an integrated delivery system.

Panel member 5: Based on method and result
Panel member 6: No concerns, among the sample 68 commericial and 19 Medicaid plans, the composite measure appears to be highly reliable, as are the component measures.

Validity

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    Panel member 1: N/A
    Panel member 2: No concerns

    Panel member 3: SMO: None

    Panel member 4: In the documentation, it does not appear as if the measure developer looked that the impact of excluding hospice patients. The results presented on exclusions focused on deliveries less than 37 weeks gestation.

    Panel member 5: None

    Panel member 6: None

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

    Submission document: Testing attachment, section 2b4.

    Panel member 1: None – the testing sample includes 68 commercial plans and 19 Medicaid plans.

    Panel member 2: No concerns

    Panel member 3: SMO: None. Data presented in Section 2b4 show wide variation in performance. The ability to identify statistically significant differences is also reflected in the high estimated signal-to-noise reliability.

    Panel member 4: None. See substantial variation across all vaccination types and all plan types.

    Panel member 5: Appears the IQRs are statistically significant (likely influenced by large sample size); however, practical difference across health plans isn't described.

    Panel member 6: None

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Testing attachment, section 2b5.

Panel member 1: N/A

Panel member 2: No response

Panel member 3: SMO: None
Panel member 4: N/A

Panel member 5: None

Panelmember 6: None

15. **Please describe any concerns you have regarding missing data.**

Submission document: Testing attachment, section 2b6.

Panel member 1: None

Panel member 2: No concerns

Panel member 3: SMO: None
Panel member 4: None. NCQA has standard processes for ensuring data are captured.

Panel member 5: None

Panelmember 6: No response

16. **Risk Adjustment**

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

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

☐ Yes ☐ No ☒ Not applicable

Panel member 4: (**process measure**)

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

Panel member 1: No risk-adjustment method was used. Social risk factors were not used because of lack of such data at health plan level, except that data from Commercial and Medicaid plans were analyzed separately, which may serve as proxy for income.

Panel member 2: No response

Panel member 3: SMO: This is a process measure and was not risk-adjusted. In theory, differences across plans could be explained in part by case mix.

Panel member 4: No response

Panel member 5: N/A

Panelmember 6: No response

VALIDITY: TESTING

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

    ☒ Face validity

    ☒ Empirical validity testing of the measure score

    ☐ N/A (score-level testing not conducted)
19. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    Panel member 1: Empirical validity testing through Pearson correlation was used for assessing construct validity for composite and component measure scores. The measure developer also assessed correlation of *prenatal immunization status* with *adult immunization satatus*, *childhood immunization status* and *immunization of adolescent combination rates*.

    Systematic assessment of face validity was conducted through different advisory panels and through public participation.

    Panel member 2: Pearson correlation for Empiric Testing was appropriate. Face Validity of Composite and Component Performance Measure Score thorough.

    Panel member 3: SMO: The developers assessed correlations between each individual measure and the overall all-or-none composite. They also compared score-level results for this measure to other HEDIS vaccine measures.

    Panel member 4: Face validity - process used does not seem to match NQF's criteria; process does not specifically address the questions that NQF wants asked
    Empirical testing – compared whether the indicators in the measure correlate with each other; also looked at performance on this measure vs. other HEDIS vaccination measures

    Panel member 5: Pearson correlation (component measures with whole and whole with other measures)

    Panelmember 6: The developer evaluated the correlation between data elements and measure score with other vaccinations and did conducted what appears to be a passive test of face validity. The developer indicated that 'face validity is systematically determined' but no description of that method was provided. The correlations were conducted at the health plan level, but this form of testing represents a lower level of validity testing, as it is not measuring whether the appropriate action occurred, or if outcomes were different for patients who met the measure.

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

    Panel member 1: Correlations among the rates within the composite measure were strong, which seem to suggest that if any plan performs well on any of the vaccination rate, it is expected to perform well on the other vaccine as well.

    Panel member 2: Coefficients of performance scores demonstrated moderate to high correlation values using the submitors thresholds.

    Panel member 3: SMO: Results indicated that each individual measure was highly correlated with the composite. In addition, there was a positive correlation between performance on this composite and other

HEDIS vaccine measures.

Panel member 4: Face validity –results provided only reflected NCQA's internal processes for measurement development

Empirical testing – generally showed strong correlations with indicators within the measure (Table 4); correlations with other HEDIS vaccination measures were generally moderate (Tables 5-6)

Panel member 5: Moderate to high correlations, with higher correlations among components with the whole.

Panelmember 6: Correlations showed generally strong associations between the measure score, measure components and other vaccinations. This means that plans that had higher rates of prenatal immunization also had higher rates of other immunizations. Face validity methodology does not appear to be systematic, so will not be factored in by this reviewer.

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    Panel member 6: (marginally so)

    ☐ **No**

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

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

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

    ☒ **Yes**

    ☐ **No**

    ☒ **Not applicable**

    Panel member 4: (data element testing was not performed)

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

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

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

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

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

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

    Panel member 1: See comments in 22 above

    Panel member 2: Similar concerns as to the "Strong" "Moderate" "Weak" thresholds I described for Adult Immun 3483 although the results for this measure demonstrated stronger correlation in the results stratified by health plans.

Panel member 3: SMO: This measure appears to measure what it purports to measure and is related to quality.

Panel member 4: Measure developer used appropriate empirical approach; strong correlations with indicators within the measure (Table 4); correlations with other HEDIS vaccination measures were generally moderate (Tables 5-6). I'm not sure why the measure developer hypothesized that vaccination success with pediatrics would translate to adults and adolescents (my experience is the engagement strategies vary across populations).

Panel member 5: Validating a measure with itself is not a strong approach. As expected the correlations dropped significantly when comparing to alternate measures (low of .16).

Panelmember 6: As noted above, the correlations showed generally strong associations between the measure score, measure components and other vaccinations. This means that plans that had higher rates of prenatal immunization also had higher rates of other immunizations. This is not strong indication of the measure's validity but it does show that plans who tend to give one type of vaccination are likely to give others.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☒ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

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

    Panel member 1: High internal consistency was demonstrated for combining the two components (flu vaccine and Tdap vaccine) through Cronbach alpha.

    Panel member 2: Cronbach's alpha statistic demonstrated high internal consistency. TEP feedback on construct is acceptable.

    Panel member 3: SMO: Each individual measure was highly correlated with the overall composite.

    Panel member 4: The analysis supports that health plans that do well on individual components of the measure also do well on the composite score.

    Panel member 5: It is difficult to assess, possibly insufficient, but it appears problematic to have such variability in which vaccines are driving the composite score across plans. It is positied that this is due to differing quality improvement efforts—can this be tested? It seems important to validate what is driving the scores.

    Panelmember 6: The Cronbach's alpha estimate for the composite components were very high in commercial and Medicaid populations (>=0.95). The standing committee should weigh in on the conceptual appropriateness of the all or none nature of this composite measure.

ADDITIONAL RECOMMENDATIONS

27. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

Panel member 1: For implementation purpose, the measure needs to clarify on the required length of continuous enrollment in the health plan (See my comment in #2 & #11).

Panel member 2: No response

Panel member 3: No response
Panel member 4: No response

Panel member 5: No response

Panelmember 6: The standing committee should weigh in on the conceptual appropriateness of the all or none nature of this composite measure as well as the overall importance of the measure.

Questions for the Committee regarding reliability:

- The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

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

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

Preliminary rating for composite construction: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

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

2a1. Reliability-Specifications

- There was some uncertaintity expressed in the summary from the methods panel regarding the population enrollment status. Since this is a Hedis measure I would expect that the definitions would be clear, however I couldn't view them in the information that I had.
- No concerns regarding logic. Calculating gestational age for the purposes of determining inclusion could make implementation challenging, particularly in a Medicaid population where eligibility/enrollment can be triggered by pregnancy.
- No concerns
- Reliability is appropriate.
- No problems noted.
- No concerns

2a2. Reliability – Testing, any concerns?

- No. The reliablity appears very good as measured by the distribution of Beta-Binomial Statistic
- No
- No
- I agree with high rating.
- Testing was done appropriately at the measure score level, and with appropriate data. However, as noted above, the vague specification of the term 'up-to-date' is a major concern. The beta-binomial method is appropriate for the individual vaccines, but since the composite is not a binomial proportion as noted above, it is not strictly appropriate for the composite measure. However, since the results are so strong, I don't see a problem. The reliability results look high, but given median plan sizes of 1,514 and 4,408 and the broad range of coverage rates, this is not surprising. Overall rating of reliability: High.
- High rate of reliability testing

2b1. Validity -Testing, any concerns

- No
- no
- No concerns
- Validity is appropriate.
- No. Face validity approach is appropriate.
- No concerns

2b4-7; 2b2-3. Threats to Validity

- Ok
- Measure appears to identify some meaningful quality differences. A process appears to be in place for continued vetting of submissions to ensure that missing data does not unduly bias outcomes.
- No concerns
- No concerns noted.
- Data used for testing are appropriate.
- No concerns here.
- Excludes women who had a preterm delivery, what is the rationale for this exclusion? Is this due to the time period for vaccination to be in the last trimester?
- Exclusions appear consistent with evidence. No groups inappropriately excluded
- Again, I am surprised by the designation of "not applicable". There should be an exploration of social risk adjustment to identify prioirty populations.
- No concerns noted.
- The exclusion of women giving birth prematurely (before 37 weeks) may be problematical: 10% - 13% of the observations are missing, and one suspects that women carrying to term may have different vaccine uptakes than women giving birth prematurely. The rationale is that women giving birth prematurely may not have had enough time to receive the vaccine, but the possibility of a premature birth seems like something that prenatal care providers should consider. As a point of reference, CDC recommends TDaP in the 3rd trimester, but flu in any semester, presumably based on time of year.
- It appears risk adjustment data is not noted--which could pose as a threat to validity, especially while consider social risk adjustment as a factor.

2c. Composite Performance Measure

- Yes
- Per submission narrative, no weighting applied to measures. Testing appears appropriate.
- I do have questions about the "all or nothing" nature of the measure. Someone with more specific expertise than I would know whether these are the "right" 2 vaccines to include in such a composite
- No concerns noted.
- Correlations between coverage fractions for pre-natal vaccines in the composite are high, and higher than with vaccines for the adult and adolescent populations, as one would expect. This supports the concept of the composite. Overall rating of validity: High
- N/A

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.

ALL data elements are in defined fields in a combination of electronic sources. Data Elements generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry).

The measure developer notes that an independent audit of all HEDIS collection and reporting processes is conducted, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. In addition to the HEDIS audit, the measure developer provides a system to allow "real-time" feedback from measure users.

Questions for the Committee:

- None

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

- Should be routinely generated during care if data are from Claims. Registry data may or may not be feasible depending on if a passive or active system. They note that DRG codes could impact data collection.
- Data is routinely generated as part of care provision; no concerns regarding collection
- No concerns
- I agree with moderate rating.
- The individual vaccine measures are currently in use in HEDIS with no apparent problems. Rating: High
- Rate of feasibility for this measure is moderate given the HEDIS collection and reporting processes involved (e.g., health plan "chart chase")

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

The developer provided a table of current and planned use

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting Quality Improvement (external benchmarking to organizations)	Quality Improvement (Internal to the specific organization) Healthcare Effectiveness Data and Information Set (HEDIS) https://www.ncqa.org/hedis/using-hedis-measures/

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

The HEDIS set is one of health care's most widely used performance improvement tools and are used by health plans and other various levels of the health care system for quality improvement initiatives. This measure was a new HEDIS measure in 2018. NCQA's standard process is to evaluate data for all new measures prior to use for public reporting, benchmarking and/or other programs.

The developer notes that the plan for implementation includes working with a multi-stakeholder advisory panels to assess the number of plans that have shown they can report the measure; whether measure results match what we expect; whether results seem indicative of true performance; and whether performance indicates an opportunity for improvement for the industry overall.

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

Feedback on the measure by those being measured or others

The measure developer noted that during a recent public comment posting which was held during the development process, measured entities supported the new measure and found it to be relevant and clearly specified.

Additional Feedback:

Questions for the Committee:

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

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

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

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

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

Improvement results New measure, therefore no data on improvement over time. Adoption of this measure has the potential to improve the immunization rates for adults.

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

Unexpected findings (positive or negative) during implementation [unexpected findings]

Potential harms None Idenified

Additional Feedback:

Questions for the Committee:

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

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

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

4a1. Use - Accountability and Transparency

- HEDIS Measure. The components of this composite are used in public health surveillance and this composite would also be useful.
- Measure is still undergoing NCQA internal review, not part of public reporting yet. Approval for public reporting anticipated. Implementation plan details limited.
- No concerns. Both are currently in use.
- I agree with passing rating.
- The individual vaccine measures are currently used effectively in HEDIS. Rating: High
- Currently, there is public reporting and use in an accountability program for this measure.

4b1. Usability – Improvement

- The results would be useful for public and clinical healthcare providers to improve prenatal vaccination rates.
- No data on improvement yet available; potential stated but not detailed.
- I do not perceive any unintended harms
- I agree with moderate rating.
- NA
- Progress toward achieving the goal of high-quality, efficient healthcare for the maternal population is demonstrated. Further, adoption of this measure has the potential to improve the immunization rates for mothers and improve overall maternal mortality in our country.

Criterion 5: Related and Competing Measures

Related or competing measures

0039 : Flu Vaccinations for Adults Ages 18 and Older

0041 : Preventive Care and Screening: Influenza Immunization

0431 : INFLUENZA VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

0680 : Percent of Residents Who Were Assessed and Appropriately Given the Seasonal Influenza Vaccine (Short Stay)

0681 : Percent of Residents Assessed and Appropriately Given the Seasonal Influenza Vaccine (Long Stay)

1659 : Influenza Immunization

Harmonization

The measure developer notes that the specifications are not harmonized. This measure specifically assesses immunizations administered during prenatal care. Other related measures assess broader populations and older adults, and do not provide information about the quality of care provided to pregnant women.

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

- This measure is specific to prenatal care.
- No additional steps given for harmonization; may not be necessary due to specificity of target sub-population.

- No concerns noted.
- NA
- It is important to note that this measure specifically assesses immunizations administered during prenatal care. Other related measures assess broader populations and older adults, and do not provide information about the quality of care provided to pregnant women, which is a significant concern, considering our increasing maternal mortality when compared to other developed nations in the world.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: Month/Day/Year

- **Of the XXX NQF members who have submitted a support/non-support choice:**
  - XX support the measure
  - YY do not support the measure

Developer Submission

Brief Measure Information

NQF #: 3484

Corresponding Measures:

De.2. Measure Title: Prenatal Immunization Status

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

De.3. Brief Description of Measure: Percentage of deliveries in the measurement period in which women received influenza and tetanus, diphtheria toxoids and acellular pertussis (Tdap) vaccinations.

1b.1. Developer Rationale: See question 1c.3.

S.4. Numerator Statement: Deliveries in which women received influenza and tetanus, diphtheria toxoids and acellular pertussis (Tdap) vaccinations.

S.6. Denominator Statement: Deliveries that occurred during the measurement period.

S.8. Denominator Exclusions: Deliveries that occurred at less than 37 weeks gestation.

Deliveries in which women were in hospice during the measurement period.

De.1. Measure Type: Composite

S.17. Data Source: Claims, Electronic Health Data, Electronic Health Records, Enrollment Data, Management Data, Other, Registry Data

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not applicable.

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

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

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

PRS_Evidence_Form.docx

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

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

No

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 3484

Measure Title: Prenatal Immunization Status

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 11/15/2019

Instructions

- ***Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.***
- ***Complete EITHER 1a.2, 1a.3 or 1a.4 as applicable for the type of measure and evidence.***
- ***For composite performance measures:***
  - ***A separate evidence form is required for each component measure unless several components were studied together.***
  - ***If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.***
- **All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- **Outcome: [3] Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.**
- **Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.**
- **Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.**
- **Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.**
- **Efficiency: [6] evidence not required for the resource use component.**
- **For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.**
- **Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.**

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Outcome: Click here to name the health outcome

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

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

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

☐ Process:

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

☐ Structure: Click here to name the structure

☒ Composite: The percentage of deliveries in the measurement year in which women had received influenza and tetanus, diphtheria and acellular pertussis (Tdap) vaccinations.

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

Pregnant women >> Tdap and influenza vaccinations during pregnancy are given >> increased resistance and prevention of diseases for mother and infant >> improved health, length, and quality of life

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

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

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

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based

on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Table 1. Tetanus, Diphtheria Toxoids and Acellular Pertussis (Tdap) Vaccine Recommendation

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Title: Prevention of Pertussis, Tetanus, and Diphtheria with Vaccines in the United States: Recommendations of the Advisory Committee on Immunization Practices (ACIP)** • **Author: Jennifer L. Liang, National Center for Immunization and Respiratory Diseases, CDC** • **Date: April 27, 2018** • **Citation: Liang J, Tiwari T, Moro P et al. Prevention of Pertussis, Tetanus, and Diphtheria with Vaccines in the United States: Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2018;67(No. 2):1-44.** • **URL: https://www.cdc.gov/mmwr/volumes/67/rr/pdfs/rr6702a1-H.pdf**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	"Health care personnel should administer a dose of Tdap during each pregnancy, irrespective of the patient's prior history of receiving the vaccine. Tdap should be administered between 27 and 36 weeks' gestation, although it may be administered at any time during pregnancy. Available data suggest that vaccinating earlier in the 27–36 week time period will maximize passive antibody transfer to the infant. Tdap may be simultaneously administered with an inactivated influenza vaccine to pregnant women."
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	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.
Provide all other grades and definitions from the evidence grading system	N/A

Grade assigned to the **recommendation** with definition of the grade	ACIP did not provide a grade for this recommendation. CDC vaccine recommendations are developed using an explicit evidence-based method based on the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach. Key factors considered in development of recommendations include balance of benefits and harms, type or quality of evidence, values and preferences of the people affected, and health economic analyses. ACIP discusses recommendations with professional organizations and holds regular meetings for experts to vote on proposed recommendations.
Provide all other grades and definitions from the recommendation grading system	N/A
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	The ACIP Pertussis Vaccines Work Group reviewed available published and unpublished data and evidence from 2004 to 2017, covering topics such as tetanus, diphtheria and pertussis disease epidemiology in the United States, decision analyses, cost-effectiveness, programmatic considerations, vaccine immunogenicity, vaccine safety, and postlicensure Tdap vaccine effectiveness. In total, they reviewed 110 studies consisting of randomized control trials and other types of studies for Tdap and Td vaccination, and approximately 77 studies on prenatal Tdap vaccination (including 16 studies on prenatal Tdap vaccine effectiveness and 5 studies on timing of Tdap vaccination during pregnancy).
Estimates of benefit and consistency across studies	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 the vaccination of women during each pregnancy with a single dose of Tdap. The Work Group reviewed data that concluded that "the strategy of preventing pertussis in newborns through the vaccination of women with Tdap during pregnancy from 27 through 36 weeks' gestation is 80%–91% effective."
What harms were identified?	The Work Group examined the risk of adverse events for each vaccine component and concluded that the Tdap vaccine is safe for administration to pregnant women. Any adverse reactions that were observed were limited to minor local reactions, including pain, erythema and swelling; no serious adverse events have been observed. Receipt of Tdap during pregnancy has not been found to be associated with an increased risk for frequency of major malformations, stillbirth, preterm birth, small for gestational age, or hypertensive disorders. 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.

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	There have been no studies published since the guideline that would significantly affect the findings.

Table 2. Influenza Vaccine Recommendation

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Title: Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices – United States, 2019-20 Influenza Season** • **Author: Lisa A. Grohskopf, Influenza Division, National Center for Immunization and Respiratory Diseases, CDC** • **Date: August 23, 2019** • **Citation: Grohskopf LA, Alyanak E, Broder KR, Walter EB, Fry AM, Jernigan DB. Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices — United States, 2019–20 Influenza Season. MMWR Recomm Rep 2019;68(No. RR-3):1–21. DOI: http://dx.doi.org/10.15585/mmwr.rr6803a1** • **URL: https://www.cdc.gov/mmwr/volumes/68/rr/pdfs/rr6803-H.pdf**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	"ACIP and the American College of Obstetricians and Gynecologists recommend that all women who are pregnant or who might be pregnant during the influenza season receive influenza vaccine. Any licensed, recommended, and age-appropriate inactivated influenza vaccine (IIV4) or recombinant influenza vaccine (RIV4) may be used. Live attenuated influenza vaccine (LAIV4) should not be used during pregnancy. Influenza vaccine can be administered at any time during pregnancy, before and during the influenza season."
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	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.
Provide all other grades and definitions from the evidence grading system	N/A
Grade assigned to the **recommendation** with definition of the grade	ACIP did not provide a grade for this recommendation. CDC vaccine recommendations are developed using an explicit evidence-based method based on the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach. Key factors considered in development of recommendations include balance of benefits and harms, type or quality of evidence, values and preferences of the people affected, and health economic analyses. ACIP discusses recommendations with professional organizations and holds regular meetings for experts to vote on proposed recommendations.

Provide all other grades and definitions from the recommendation grading system	N/A
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	The ACIP Influenza Work Group reviewed available data and evidence from 1979 to 2018 on immunogenicity, efficiency, effectiveness and safety of influenza vaccines. They also convene twice monthly to review "influenza surveillance, vaccine effectiveness and safety, vaccine coverage, program feasibility, cost-effectiveness, and vaccine supply" in order to provide annual recommendations for the use of influenza vaccines for the prevention and control of influenza. In total, the Work Group reviewed approximately 285 studies on the immunogenicity, efficacy and effectiveness of IIV, RIV and LAIV and 120 studies on influenza vaccine safety. This review includes approximately 12 studies specifically assessing immunogenicity, efficacy and effectiveness of influenza vaccine for pregnant women; and 46 studies assessing influenza vaccine safety for pregnant women. These studies consist of randomized control trials, case control studies and observational studies, among others.
Estimates of benefit and consistency across studies	The Work Group reviewed vaccine safety—compared to the morbidity and mortality risks associated with influenza—and concluded that all persons aged ≥6 months without contraindications are recommended to receive routine influenza vaccinations. Vaccine type, product, and dose recommendations are based on age and pregnancy status. This includes vaccination of pregnant women with a licensed, recommended, and age-appropriate IIV or RIV4.
What harms were identified?	The Work Group examined the risk of adverse events for each vaccine type and concluded that the influenza vaccine is safe for routine administration, including administration of IIV4 and RIV4 to pregnant women (LAIV4s are contraindicated for pregnant women). ACIP has identified severe allergic reactions as a contraindication to all influenza vaccine types. ACIP also notes that: • Influenza vaccines have been administered to pregnant women for more than five decades, and overall have a reassuring safety profile. The vast majority of published data and clinical experience involve use of IIVs, which have been available for the longest period of time, and which have been recommended for use for some populations of pregnant women since the early 1960s. Data are more limited for RIV4 (which has only been available since 2013). • Substantial data have accumulated which do not indicate fetal harm (including death, spontaneous abortion or congenital malformations) associated with IIVs administered during pregnancy. • Assessments of association between IIV and preterm birth and small for gestational age infants have yielded inconsistent results, with most studies reporting no association or a protective effect against these outcomes.

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	There have been no studies published since the guideline that would significantly affect the findings.

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

1b. Performance Gap

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

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

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

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

See question 1c.3.

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

The following data are extracted from HEDIS data collection reflecting the most recent year of measurement (2018) for this measure. Performance data is summarized at the health plan level and summarized by mean performance and performance at 10th, 25th, 50th, 75th, and 90th percentile. We also calculated the interquartile range (IQR), which can be interpreted as the difference between the 25th?and 75th?percentile. Data is stratified by product line (i.e. commercial and Medicaid).

The following data demonstrate the variation in the rate of prenatal immunization across health plans. For the indicator assessing receipt of influenza vaccination among pregnant women, there was a 12 point difference between plans in the 25th percentile and plans in the 75th percentile for commercial plans and 11 points for Medicaid plans. For the indicator assessing receipt of Tdap vaccination among pregnant women, there was a 17 point difference between plans in the 25th percentile and plans in the 75th percentile for commercial plans and 16 points for Medicaid plans. These gaps in performance underscore the opportunity for improvement.

Prenatal Immunization Status: Both Vaccines

Commercial, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

33.1 | 18.4 | 26.6 | 33.6 | 39.0 | 44.5 | 12.4

Medicaid, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

16.7 | 8.1 | 12.2 | 17.0 | 19.6 | 25.3 | 7.4

Prenatal Immunization Status: Influenza

Commercial, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

40.5 | 27.3 | 33.3 | 40.7 | 45.5 | 52.4 | 12.2

Medicaid, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

23.8 | 13.1 | 17.2 | 23.5 | 28.0 | 32.2 | 10.8

Prenatal Immunization Status: Tdap

Commercial, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

62.7 | 44.5 | 55.2 | 65.4 | 72.2 | 77.3 | 17.0

Medicaid, 2018

Mean | 10th | 25th | 50th | 75th | 90th | Interquartile Range

40.4 | 27.2 | 33.3 | 40.6 | 48.8 | 56.3 | 15.5

The HEDIS performance data reflect the most recent year of measurement for this measure. Below is a description of the number of health plans that reported this measure and the median denominator for the measure (stratified by commercial and Medicaid).

Commercial, 2018

N Plans | Median Denominator Size

68 | 1,374

Medicaid, 2018

N Plans | Median Denominator Size

19 | 3,800

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

Studies have found that about half of women do not receive the influenza vaccine and/or the Tdap vaccine during pregnancy. The CDC conducted an internet panel survey and found that 50 percent of women who were pregnant any time between October 2014 and January 2015 reported receiving the influenza vaccine after July 2014 (Ding et al 2015). Twenty percent of surveyed women said their provider did not recommend or offer the vaccine. In the Pregnancy Risk Assessment Monitoring System survey (PRAMS), 53 percent of women who had a live birth in 2011 reported receiving the Tdap vaccine during pregnancy, although 20 percent of the women surveyed did not know their immunization status (Ahluwalia et al 2015).

A study from the 2013–2014 influenza season using patient-reported and vital records data indicated that 41 percent of pregnant women received the influenza vaccine during pregnancy (Kerr et al 2016). A separate study that matched prenatal care data from patient vital records and data from the Minnesota state

immunization registry found that 46 percent of women who had given birth in Minnesota from 2013–2014 had received the influenza immunization during pregnancy, and 58 percent received the Tdap vaccine during pregnancy (Barber et al 2017). Among pregnant women who received the Tdap vaccine, 86 percent received it during the optimal timing of 27–36 weeks gestation (Barber et al 2017).

In a 2014 study using chart review data from a private physician office and a resident clinic, 66 percent of women received the Tdap vaccine during pregnancy; of these, 91 percent received it during the recommended time frame (Ravin 2016). Few women in this study declined the vaccine, leading the researchers to conclude that higher immunization rates would likely be achieved if vaccines were offered more often (Ravin 2016). Likewise, reminder systems and standing orders that allow members of the health care team other than the attending provider to assess vaccination status and administer vaccines can help to ensure wider vaccination coverage (Ding et al 2015).

Citations:

Ahluwalia, I., H. Ding, D. D'Angelo, K. Shealy, J. Singleton, J. Liang, K. Rosenberg. 2015. "Tetanus, Diphtheria, Pertussis Immunization Coverage Before, During, and After Pregnancy—16 States and New York City, 2011." MMWR Morb Mortal Wkly Rep. 64(19);522–6.

Barber, A., M.H. Muscoplat, A. Fedorowicz. 2017. "Coverage with Tetanus, Diphtheria, and Acellular Pertussis Vaccine and Influenza Vaccine Among Pregnant Women—Minnesota, March 2013–December 2014." MMWR Morb Mortal Wkly Rep. 66:56–59. DOI: http://dx.doi.org/10.15585/mmwr.mm6602a4.

Ding, H., C.L. Black, S. Ball, et al. 2015. "Influenza immunization coverage among pregnant women—United States, 2014–15 influenza season." MMWR Morb Mortal Wkly Rep. 64(36);1000–5.

Kerr, S., C.M. Van Bennekom, A.A. Mitchell. 2016. "Influenza Immunization Coverage During Pregnancy—Selected Sites, United States, 2005–06 through 2013–14 Influenza Vaccine Seasons." MMWR Morb Mortal Wkly Rep. 65:1370–1373. DOI: http://dx.doi.org/10.15585/mmwr.mm6548a3.

Ravin, A., J. Koerner, A. Forinash, A. Bergin, K. March, C. Miller. 2016. "Rates of Adherence to Tdap Immunization Guidelines in Pregnancy." Obstetrics and Gynecology. doi: 10.1097/01.AOG.0000483449.15059.7d.

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

HEDIS data are stratified by type of insurance (e.g. Commercial, Medicaid, Medicare). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities, if the data are available to a plan. 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.

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

Prenatal immunization rates vary based on patient race, ethnicity, age, insurance status and adequacy of prenatal care. A CDC panel survey of women who were pregnant any time between October 2014 and January 2015 found that 39 percent of non-Hispanic Black women had received the influenza immunization after July

2014, compared with 52 percent of non-Hispanic White women (Ding et al 2015). 62 percent of pregnant women 35 or older received the influenza vaccine, compared with 50 percent of pregnant women 25–34 and 44 percent of pregnant women 18–24 (Ding et al 2015). 57 percent of women with private health insurance received the influenza vaccine, compared with 40 percent of women with public health insurance and 3 percent of women with no insurance (Ding et al 2015).

An analysis from the 2012–2015 National Health Interview Survey (NHIS) found that non-Caucasian ethnic groups, African Americans, women without a usual source of health care and women with higher alcohol consumption were less likely to receive an influenza vaccine during pregnancy (Chan et al 2017). The analysis also found that higher education and income levels were associated with higher influenza vaccination rates (Chan et al 2017).

PRAMS survey data from the 2009–2010 influenza season revealed that influenza vaccination coverage among women with live births was 51 percent for non-Hispanic White women, compared with 30 percent for non-Hispanic Black women and 42 percent for Hispanic women (Ahluwalia et al 2014). Data from the PRAMS survey for Tdap vaccination indicate that vaccination coverage was lower for non-Hispanic Black women, those with Medicaid insurance and those starting prenatal care after the first trimester of pregnancy (Ahluwalia et al 2015).

Citations:

Ahluwalia, I., H. Ding, D. D'Angelo, K. Shealy, J. Singleton, J. Liang, K. Rosenberg. 2015. "Tetanus, Diphtheria, Pertussis Immunization Coverage Before, During, and After Pregnancy—16 States and New York City, 2011." MMWR Morb Mortal Wkly Rep. 64(19);522–6.

Ahluwalia, I., H. Ding, L. Harrison, D. D'Angelo, J. Singleton, C. Bridges. 2014. "Disparities in Influenza Vaccination Coverage among Women with Live-Born Infants: PRAMS Surveillance during the 2009–2010 Influenza Season." Public Health Reports. 129(5):408–16.

Chan, H., J. Chang., S.R. Erickson, C. Wang. 2017. "Influenza Vaccination Among Pregnant Women: Exploratory Analysis From The 2012-2015 National Health Interview Survey." Value in Health. 20: A797.

Ding, H., C.L. Black, S. Ball, et al. 2015. "Influenza immunization coverage among pregnant women—United States, 2014–15 influenza season." MMWR Morb Mortal Wkly Rep. 64(36);1000–5.

1c. Composite Quality Construct and Rationale

1c.1. A composite performance measure is a combination of two or more component measures, each of which individually reflects quality of care, into a single performance measure with a single score.

For purposes of NQF measure submission, evaluation, and endorsement, the following will be considered composites:

- Measures with two or more individual performance measure scores combined into one score for an accountable entity.
- Measures with two or more individual component measures assessed separately for each patient and then aggregated into one score for an accountable entity:
    - all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient);

1c.1. Please identify the composite measure construction: all-or-none measures (e.g., all essential care processes received, or outcomes experienced, by each patient)

1c.2. Describe the quality construct, including:

- the overall area of quality
- included component measures and
- the relationship of the component measures to the overall composite and to each other.

This measure assesses the provision of critical immunizations for prenatal women per clinical guidelines. The intent of the measure is to improve primary prevention of vaccine-preventable diseases for the mother and baby, including influenza and tetanus, diphtheria and pertussis (whooping cough). The measure calculates a rate for each specific vaccine (influenza and Tdap) and an all-or-nothing composite rate assessing receipt of both vaccines.

1c.3. Describe the rationale for constructing a composite measure, including how the composite provides a distinctive or additive value over the component measures individually.

The combination rate provides an overview of whether prenatal women received both recommended vaccines. The individual vaccine component rates are included to provide information on which prenatal vaccinations (Tdap and influenza) are being provided to members as recommended.

1c.4. Describe how the aggregation and weighting of the component measures are consistent with the stated quality construct and rationale.

The components are weighted equally in an all-or-none composite to assess compliance with immunization guidelines for prenatal women.

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

NATIONAL QUALITY FORUM—Composite Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Composite Measure Title: Prenatal Immunization Status
Date of Submission: 8/1/2019

Composite Construction:

☐Two or more individual performance measure scores combined into one score

☒ All-or-none measures (e.g., all essential care processes received or outcomes experienced by each patient)

Instructions: Please contact NQF staff before you begin.

- **If a component measure is submitted as an individual performance measure, the non-composite measure testing form must also be completed and attached to the individual measure submission.**
- **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.**
- **Sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For composites with outcome and resource use measures, section 2b3 also must be completed.**
- **If specified for multiple data sources/sets of specificaitons (e.g., claims and EHRs), section 2b5 also must be completed.**
- **Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) and composites (2c) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). *Contact NQF staff if more pages are needed.***
- **Contact NQF staff regarding questions. Check for resources at 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 version 7.1 of the Measure Testing Attachment. and the 2017 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.

2a2. Reliability testing [10] 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 [11] 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 evidenceand are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]

AND

If patient preference (e.g., informed decisionmaking) 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). [13]

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**

OR

- **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 nonresponders) 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)*

Notes

10. 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).

11. 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.

12. 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.

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

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

15. 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.

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

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

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

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

☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
other:	other:

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

1.3. What are the dates of the data used in testing? 01/01/2018-12/31/2018

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

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

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

This measure assesses whether prenatal women enrolled in commercial and Medicaid health plans received influenza and Tdap vaccines per clinical guidelines. The denominator is specified as the number of deliveries during the measurement year. There is a rate for each individual vaccine and a combination rate assessing receipt of both vaccines. The intended use of the measure is to assess the quality of care in health plans in ensuring their prenatal population receives important vaccines. As required by the intended level of accountability, we assessed data from all health plans reporting the HEDIS measure to NCQA in 2018 to assess scientific acceptability, usability and feasibility.

Data were calculated from all commercial and Medicaid health plans submitting data to NCQA for this HEDIS measure to assess measure score reliability, construct validity, distribution of exclusions and average percentage of members that were excluded, meaningful differences in performance, missing data and components of the composite. Data came from 68 commercial health plans and 19 Medicaid health plans that were geographically diverse and varied in size. Data from administrative claims, electronic health records, registries, health information exchanges and case management systems were eligible for use in the measure in line with NCQA's HEDIS Electronic Clinical Data Systems reporting method. The plans submitting HEDIS data used a range of data sources: administrative claims, immunization registry, electronic health record and case management data.

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

- The Pregnancy Health Measurement Advisory Panel included 9 experts in maternal care, immunizations and measures development, as well as clinician, health-plan and state/federal representatives.
- The Technical Measurement Advisory Panel includes 12 members, including representation by health plan methodologists, clinicians, HEDIS auditors and state/federal users of measures.
- NCQA's Committee on Performance Measurement (CPM) oversees measures used in NCQA programs and includes representation by purchasers, consumers, health plans, health care providers, and policy makers. This panel is composed of 21 independent members that reflect the diversity of constituencies that performance measurement serves. The CPM's recommendations are reviewed and approved by NCQA's Board of Directors.

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

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

Table 1. Median eligible population for *Prenatal Immunization Status* by plan type, 2018

Plan Type	Number of Plans	Median number of eligible patients per plan
Commercial	68	1,514
Medicaid	19	4,408

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

The 2018 HEDIS data were used to assess measure score reliability, construct validity, distribution of exclusions and average percentage of members that were excluded, meaningful differences in performance, missing data and components of the composite, as described above. For empirical validity testing, NCQA explored whether the composite and component measure rates were correlated with other relevant HEDIS measures that the plans reported in 2018. Validity was also demonstrated through a systematic assessment of face validity.

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

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

_____

2a2. RELIABILITY TESTING

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

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

***Note**: Current guidance for composite measure evaluation states that reliability must be demonstrated for the composite performance measure score.*

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

2a2.2. 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*)

Reliability Testing of Performance Measure Score: We used the Beta-binomial model[1] to assess how well one can confidently distinguish the performance of one accountable entity from another. Conceptually, the Beta-binomial model is the ratio of signal to noise. The signal is the proportion of the variability in measured performance that can be explained by real differences in performance. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS measures. Reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (i.e., noise), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities). The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

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

Table 2. Distribution of Beta-Binomial Statistics For Each Measure Rate, Commercial Plans - 2018

Rate	Overall Reliability	Min	10th	25th	50th	75th	90th	Max
Receipt of both Vaccines	0.993	0.789	0.931	0.969	0.988	0.995	0.997	0.999
Influenza	0.993	0.796	0.931	0.965	0.986	0.995	0.997	0.999
Tdap	0.995	0.832	0.953	0.977	0.991	0.997	0.998	1.000

Table 3. Distribution of Beta-Binomial Statistics For Each Measure Rate, Medicaid Plans - 2018

Rate	Overall Reliability	Min	10th	25th	50th	75th	90th	Max
Receipt of both Vaccines	0.996	0.684	0.960	0.987	0.994	0.998	0.999	0.999
Influenza	0.996	0.730	0.917	0.987	0.994	0.998	0.999	0.999
Tdap	0.997	0.805	0.912	0.982	0.995	0.998	0.999	0.999

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

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

The values for the overall beta-binomial statistic across all product lines and measure rates are all greater than 0.7, indicating the measure has very good reliability. The distribution of health plan level-reliability on this measure shows that all health plans (across all product lines) are above the threshold of 0.7 except the Medicaid minimum for the combination rate, which is close to 0.7. Good reliability is demonstrated since most variance is due to signal and not to noise.

_____

2b1. VALIDITY TESTING

Note*: Current guidance for composite measure evaluation states that validity should be demonstrated for the composite performance measure score. If not feasible for initial endorsement, acceptable alternatives include assessment of content or face validity of the composite OR demonstration of validity for each component. Empirical validity testing of the composite measure score is expected by the time of endorsement maintenance.*

2b1.1. What level of validity testing was conducted?

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

☒ **Composite performance measure score**

☒ **Empirical validity testing**

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

☒ **Validity testing for component measures** *(check all that apply)*

Note*: applies to ALL component measures, unless already endorsed or are being submitted for individual endorsement.*

☐ **Endorsed (or submitted) as individual performance measures**

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

☒ **Empirical validity testing of the component measure score(s)**

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

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

Empiric Validity Testing of Composite and Component Performance Measure Score

We empirically evaluated composite measure score validity and component measure score validity for commercial plans using the analysis described below. We were unable to conduct these analyses with Medicaid plans due to a smaller number of plans in our sample.

Empiric validity of the results were assessed using Pearson correlation to demonstrate construct validity. This test estimates the strength of the linear association between two variables; the magnitude of correlation ranges from -1 to +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated

with decreasing values of the second variable. Coefficients with absolute values of less than 0.3 are generally considered indicative of weak associations, whereas absolute values of 0.3 or higher denote moderate to strong associations. 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 it is unlikely that a non-zero coefficient was observed due to chance alone.

For commercial and Medicaid plans, we explored whether the indicators within this measure were correlated with each other. We hypothesized that health plans that perform well on one component vaccine should perform well on the other vaccine. All of the measure rates represent an underlying quality construct of administering recommended vaccines to prenatal women.

We also assessed correlation of *Prenatal Immunization Status* with other relevant HEDIS measures that they reported in 2018. Specifically, we explored whether *Prenatal Immunization Status* measure rates were correlated with the HEDIS *Adult Immunization Status* measure rates. We also explored whether the *Prenatal Immunization Status* rates were correlated with the HEDIS *Childhood Immunization Status* and *Immunization for Adolescents* combination rates. *Childhood Immunization Status* assesses whether children received 10 recommended vaccines by their second birthday, while *Immunizations for Adolescents* assesses whether adolescents received three recommended vaccines by their 13th birthday. We hypothesized that health plans that achieve high rates on these should also perform well on the *Prenatal Immunization Status* measure.

Systematic Assessment of Face Validity of Composite and Component Performance Measure Score

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

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

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. At this step, face validity is systematically determined by the CPM, which uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment. For this measure, the CPM voted to approve moving the proposed measure forward to public comment (15 CPM members approved, 0 members opposed and 0 abstained).

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

measures. For this measure, the CPM voted to approve the measure for HEDIS health plan reporting (16 CPM members approved, 0 members opposed and 0 abstained). The Board of Directors approved the measure.

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

Table 4. Health-Plan Level Pearson Correlation Coefficients Among Prenatal Immunization Measure Performance Scores Within Measure – 2018

	Prenatal Immunization Status	**Receipt of both vaccines**	**Influenza**	**Tdap**
Commercial Plans	Receipt of both vaccines	1.00*	0.97*	0.86*
	Influenza	0.97*	1.00*	0.75*
	Tdap	0.86*	0.75*	1.00*
Medicaid Plans	Receipt of both vaccines	1.00*	0.98*	0.88*
	Influenza	0.98*	1.00*	0.80*
	Tdap	0.88*	0.80*	1.00*

*significant at p<0.05

Table 5a. Health-Plan Level Pearson Correlation Coefficients Among Prenatal Immunization Status and Adult Immunization Status Measure Performance Scores in Commercial Plans – 2018

	Adult Immunization Status			
Prenatal Immunization Status	Composite	Influenza	Td/Tdap	Herpes Zoster
Receipt of both vaccines	0.78*	0.75*	0.74*	0.61*
Influenza	0.79*	0.79*	0.74*	0.61*
Tdap	0.58*	0.54*	0.59*	0.40*

**significant at p<0.05*

Table 5b. Health-Plan Level Pearson Correlation Coefficients Among Prenatal Immunization Status and Adult Immunization Status Measure Performance Scores in Medicaid Plans – 2018

	Adult Immunization Status			
Prenatal Immunization Status	Composite	Influenza	Td/Tdap	Herpes Zoster
Receipt of both vaccines	0.89*	0.85*	0.83*	0.77*
Influenza	0.91*	0.85*	0.87*	0.78*
Tdap	0.66*	0.67*	0.60*	0.54*

**significant at p<0.05*

Table 6a. Health-Plan Level Pearson Correlation Coefficients Among Prenatal, Childhood and Adolescent Immunization Measure Performance Scores in Commercial Plans – 2018

	Childhood and Adolsecent Immunization Status

Prenatal Immunization Status	Receipt of all childhoold vaccines (DTaP, Hep A, Hep B, HiB, Influenza, IPV, MMR, Pneumococcal, Rotavirus, and VZV) by age 2	Receipt of all adolescent vaccines (Tdap, meningococcal and HPV) by age 13
Receipt of both vaccines	0.42*	0.57*
Influenza	0.45*	0.59*
Tdap	0.38*	0.47*

**significant at p<0.05*

Table 6b. Health-Plan Level Pearson Correlation Coefficients Among Prenatal, Childhood and Adolescent Immunization Measure Performance Scores in Medicaid Plans – 2018

	Childhood and Adolsecent Immunization Status	
Prenatal Immunization Status	Receipt of all childhoold vaccines (DTaP, Hep A, Hep B, HiB, Influenza, IPV, MMR, Pneumococcal, Rotavirus, and VZV) by age 2	Receipt of all adolescent vaccines (Tdap, meningococcal and HPV) by age 13
Receipt of both vaccines	0.68*	0.49*
Influenza	0.62*	0.52*
Tdap	0.57*	0.16

**significant at p<0.05*

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

Composite and Component Measure Score Validity

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

Correlations were strong among the rates within the measure. The results suggest that plans that perform well on one vaccine are likely to perform well on the other vaccine.

Beyond the within-measure correlations, we saw a moderate/strong relationship with benchmarks on other measures of quality for commercial and Medicaid plans.

- The correlation between the prenatal immunization and the adult immunization composite and component measure rates was mostly strong.
- The correlation between the prenatal immunization measure rates with the childhood and adolescent immunization combination measure rates was moderate.

2b2. EXCLUSIONS ANALYSIS

Note: *Applies to the composite performance measure, as well all component measures unless they are already endorsed or are being submitted for individual endorsement.*

NA ☐ no exclusions — *skip to section 2b4*

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

The measure assesses receipt of vaccines during the prenatal period. The initial population are deliveries during the measurement period in which women were continuously enrolled at least 28 days prior to delivery through the delivery date. In order to appropriately assign accountability for vaccine provision, the measure removes women who delivered prior to 37 gestational weeks, which is prior to optimal timing for Tdap vaccination, as these women may not have had an opportunity to receive vaccines. We assessed the distribution of the initial population and exclusions for this measure and the average percentage of members that were excluded from the initial population.

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

Table 7. Distribution of initial population and exclusions

Plan Type	No. of plans	Measure Component	Mean	Min	10th	25th	50th	75th	90th	Max
Commercial	68	Initial Population	3,195	72	287	588	1,514	3,794	5,912	49,341
		Exclusions	483	5	24	64	132	372	633	15,548
Medicaid	19	Initial Population	7,518	96	702	2,290	4,408	9,022	16,284	39,365
		Exclusions	997	8	74	244	573	1,174	1,865	5,863

Table 8. Percentage of members excluded from the initial population

Plan Type	No. of plans	Mean
Commercial	68	10%
Medicaid	19	13%

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

Removing exclusions reduced the initial population, on average, by 10% for commercial plans and by 13% for Medicaid plans. This is about what we would expect, given that research shows that the average percentage of pregnancies that end in a delivery prior to 37 gestational weeks is about 10%.[2]

[2] National Vital Statistics Reports, Births, 2014. https://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_12.pdf

Based on these findings, our expert panels recommended specifying the gestational age exclusion in the measure because women who delivered prior to 37 gestational weeks may not have had an opportunity to receive the Tdap vaccine, which is recommended to occur ideally between 27-36 gestational weeks.

_____

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
Note: *Applies to all outcome or resource use component measures, unless already endorsed or are being submitted for individual endorsement.*
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used? *(check all that apply)*

☐ **Endorsed (or submitted) as individual performance measures**

☐ **No risk adjustment or stratification**

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

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

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

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

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

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

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

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

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

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

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

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

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

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

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

2b3.9. Results of Risk Stratification Analysis:

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

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

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

Note: *Applies to the composite performance measure.*

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

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

To determine if this difference is statistically significant, NCQA calculated an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than 0.05, then the two plans' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans, one plan in the 25th percentile and another plan in the 75th percentile of performance. We used these two plans as examples of measures entities. However, the method can be used for comparison of any two measured entities.

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

Table 9. Variation in Performance Across Commercial Plans - 2018

	No. of plans	Mean denominator	Mean rate (%)	Min	10th	25th	50th	75th	90th	Max	IQR	p-value
Both vaccines	68	2,712	33.1	7.4	18.4	26.6	33.6	39.0	44.5	60.1	12.4	<0.001
Influenza	68	2,712	40.5	18.6	27.3	33.3	40.7	45.5	52.4	80.5	12.2	<0.001
Tdap	68	2,712	62.7	16.4	44.5	55.2	65.4	72.2	77.3	87.8	17.0	<0.001

IQR: Interquartile Range

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

Table 10. Variation in Performance Across Medicaid Plans - 2018

	No. of plans	Mean denominator	Mean rate (%)	Min	10th	25th	50th	75th	90th	Max	IQR	p-value
Both vaccines	19	6,521	16.7	0.9	8.1	12.2	17.0	19.6	25.3	39.8	7.4	<0.001
Influenza	19	6,521	23.8	8.7	13.1	17.2	23.5	28.0	32.2	54.5	10.8	<0.001
Tdap	19	6,521	40.4	14.8	27.2	33.3	40.6	48.8	56.3	59.1	15.5	<0.001

IQR: Interquartile Range

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

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

The IQRs ranged from 12%-17% across the commercial plan rates and from 7%-16% across the Medicaid plan rates. For example, in commercial plans, the IQR for the influenza rate was 12%. This gap represents an average of 331 additional prenatal women receiving the influenza vaccine in high-performing plans compared to low-performing plans.

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

Note: *Applies to all component measures, unless already endorsed or are being submitted for individual endorsement.*

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

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

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

statistical analysis was used)

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

Note: *Applies to the overall composite measure.*

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

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

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

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

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the*

selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data)

All of the commercial and Medicaid health plans that reported 2018 HEDIS data for this measure reported valid rates as determined by NCQA-certified auditors through the process described above. This means that auditors did not find any missing data sources for any of the health plan data submissions and determined that none of the rates were materially biased.

2c. EMPIRICAL ANALYSIS TO SUPPORT COMPOSITE CONSTRUCTION APPROACH

Note: *If empirical analyses do not provide adequate results—or are not conducted—justification must be provided and accepted in order to meet the must-pass criterion of Scientific Acceptability of Measure Properties. Each of the following questions has instructions if there is no empirical analysis.*

For this measure, which assesses the receipt of two vaccines during the prenatal period, our advisory panels concluded there was no conceptual basis for applying different weights to the different vaccines.

2d1. Empirical analysis demonstrating that the component measures fit the quality construct, add value to the overall composite, and achieve the object of parsimony to the extent possible.

2d1.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

Internal Consistency: We tested the construction of the combination rate within this measure for commercial and Medicaid health plans using Cronbach's alpha statistic (internal consistency coefficients) to measure the extent to which the components (i.e., individual vaccine rates) represent a single quality construct.

2d1.2. What were the statistical results obtained from the analysis of the components? (e.g., *correlations, contribution of each component to the composite score, etc*.; *if no empirical analysis, identify the components that were considered and the pros and cons of each*)

Internal Consistency Results: Table 11 shows the results of the calculation of the Cronbach's alpha for the combination rate in commercial and Medicaid plans. The Cronbach's alpha statistic was 0.948 and 0.958 for commercial and Medicaid plans, respectively.

Table 11. Cronbach's Alpha for Combination Rate, 2018

Plan Type	Cronbach's alpha
Commercial	0.948
Medicaid	0.958

2d1.3. What is your interpretation of the results in terms of demonstrating that the components included in the composite are consistent with the described quality construct and add value to the overall composite? (i.*e., what do the results mean in terms of supporting inclusion of the components; if no empirical analysis, provide rationale for the components that were selected)*

Our results indicated good internal consistency within the combination rate across commercial and Medicaid plans.

2d2. Empirical analysis demonstrating that the aggregations and weighting rules are consistent with the quality construct and achieve the objective of simplicity to the extent possible

2d2.1 Describe the method used (*describe the steps—do not just name a method; what statistical analysis was used; if no empirical analysis, provide justification)*

We used feedback from the Pregnancy Health Measurement Advisory Panel, the Technical Measurement Advisory Panel and the Committee on Performance Measurement to determine the appropriate construction for the measure.

2d2.2. What were the statistical results obtained from the analysis of the aggregation and weighting rules? (e.g., *results of sensitivity analysis of effect of different aggregations and/or weighting rules; if no empirical analysis, identify the aggregation and weighting rules that were considered and the pros and cons of each*)

Our panels concluded there was no conceptual basis for differential weighting between the two vaccines.

2d2.3. What is your interpretation of the results in terms of demonstrating the aggregation and weighting rules are consistent with the described quality construct? (i.*e., what do the results mean in terms of supporting the selected rules for aggregation and weighting; if no empirical analysis, provide rationale for the selected rules for aggregation and weighting*)

Our panels concluded there was no conceptual basis for differential weighting between the two vaccines.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

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

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

Not applicable.

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

This is not an eMeasure **Attachment:**

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

Attachment: 3484_PRS_Value_Sets_Fall_2019-637093372926667747.xlsx

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

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

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

Not an instrument-based measure

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

No

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

N/A

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

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

Deliveries in which women received influenza and tetanus, diphtheria toxoids and acellular pertussis (Tdap) vaccinations.

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

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

Deliveries during the measurement period in which women received influenza and tetanus, diphtheria toxoids and acellular pertussis (Tdap) vaccinations. Three numerators are reported:

Numerator 1: Deliveries where members received an influenza vaccine on or between July 1 of the year prior to the measurement period and the delivery date; or deliveries where members had an influenza virus vaccine adverse reaction any time during or before the Measurement Period.

Numerator 2: Deliveries where members received at least one Tdap vaccine during the pregnancy (including the delivery date); or deliveries where members had an anaphylactic reaction to Tdap or Td vaccine or its components any time during or before the Measurement Period or encephalopathy due to Td or Tdap vaccination (post-tetanus vaccination encephalitis, post-diphtheria vaccination encephalitis, post-pertussis vaccination encephalitis) any time during or before the Measurement Period.

Numerator 3: Deliveries that met criteria for both Numerator 1 and Numerator 2.

See attached code value sets.

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

Deliveries that occurred during the measurement period.

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

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

Deliveries that occurred during the measurement period.

Note: women who had multiple deliveries during the measurement period count multiple times.

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

Deliveries that occurred at less than 37 weeks gestation.

Deliveries in which women were in hospice during the measurement period.

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

Exclude deliveries that occurred at 37 weeks of gestation or less.

Exclude deliveries where the woman was in hospice or using hospice services during the measurement period.

See attached code value sets.

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

Not applicable.

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

Step 1: Determine the eligible population. Identify all deliveries during the measurement period (January 1 – December 31) in which the patient was continuously enrolled from 28 days prior to delivery through the delivery date.

Step 2: Determine the denominator by excluding deliveries that occurred at less than 37 gestational weeks or where women were in hospice or using hospice services during the measurement period.

Step 3: Determine the numerators:

-Numerator 1: deliveries where members received an influenza vaccine on or between July 1 of the year prior to the measurement period and the delivery date; or deliveries where members had an influenza virus vaccine adverse reaction any time during or before the Measurement Period.

-Numerator 2: Deliveries where members received at least one Tdap vaccine during the pregnancy (including the delivery date); or deliveries where members had an aanaphylactic reaction to Tdap or Td vaccine or its components any time during or before the Measurement Period or encephalopathy due to Td or Tdap vaccination (post-tetanus vaccination encephalitis, post-diphtheria vaccination encephalitis, post-pertussis vaccination encephalitis) any time during or before the Measurement Period.

-Numerator 3: Deliveries in which criteria was met for both Numerator 1 and Numerator 2.

Step 4: Calculate three measure rates:

-Numerator 1 / Denominator

-Numerator 2 / Denominator

-Numerator 3 / Denominator

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

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

Not applicable.

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

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

Not applicable.

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

If other, please describe in S.18.

Claims, Electronic Health Data, Electronic Health Records, Enrollment Data, Management Data, Other, Registry Data

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

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

This measure is specified for administrative claims, electronic health record, registry, health information exchange or case management data collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from Health Management Organizations and Preferred Provider Organizations via NCQA's online data submission system.

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

No data collection instrument provided

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

Health Plan

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

Outpatient Services

If other:

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

The components are weighted equally in an all-or-none composite to assess compliance with immunization guidelines for prenatal women.

2. Validity – See attached Measure Testing Submission Form

PRS_Composite_Testing_Form.docx

2.1 For maintenance of endorsement

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

No

2.2 For maintenance of endorsement

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

No

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in a combination of electronic sources

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

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

Attachment:

3c. Data Collection Strategy

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

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

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

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

The HEDIS Compliance Audit addresses the following functions:

1)Information practices and control procedures

2)Sampling methods and procedures

3)Data integrity

4)Compliance with HEDIS specifications

5)Analytic file production

6)Reporting and documentation

In addition to the HEDIS audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system, NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system informs both annual updates to the measures as well as routine re-evaluation of measures. These processes include updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence.

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

Broad public use and dissemination of these measures are encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior

written consent of NCQA. As used herein, "commercial use" refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting Quality Improvement (external benchmarking to organizations)	Quality Improvement (Internal to the specific organization) Healthcare Effectiveness Data and Information Set (HEDIS) https://www.ncqa.org/hedis/using-hedis-measures/

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

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

HEDIS: The Healthcare Effectiveness Data and Information Set (HEDIS) is one of health care's most widely used performance improvement tools.190 million people are enrolled in health plans across the nation that report HEDIS results. HEDIS measures are used by health plans and other various levels of the health care system for quality improvement initiatives.

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

This measure was a new HEDIS measure in 2018. NCQA's standard process is to evaluate data for all new measures prior to use for public reporting, benchmarking and/or other programs.

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

As part of new measure evaluation, NCQA works with multi-stakeholder advisory panels to assess the number of plans that have shown they can report the measure; whether measure results match what we expect; whether results seem indicative of true performance; and whether performance indicates an opportunity for improvement for the industry overall. Because this measure uses the newer HEDIS Electronic Clinical Data Systems Reporting Method, NCQA's timeline and plan are to assess these issues after each year of the

measure's reporting. We anticipate that the measure will be approved for public reporting and eligible for use in programs within the next several years, but this is pending our continued assessment.

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

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

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

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

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

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

Describe how feedback was obtained.

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

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

During a recent public comment posting held during the measure development process, most of the comments from measured entities supported the new measure. In general, respondents found the measures to be relevant and clearly specified.

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

This measure has been deemed a priority measure by NCQA and other entities such as the Centers for Disease Control and Prevention, the federal National Vaccine Program Office, American College of Obstetricians and Gynecologists and the American Immunization Registry Association. During a recent public comment posting conducted during the measure development process, commenters were supportive of the measure and specifically highlighted the need for measures assessing immunizations in pregnant women. Commenters noted that many pregnant women still do not receive these important vaccines, despite Advisory Committee on Immunization Practices recommendations and national efforts to improve prenatal immunization rates in the US.

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

During measure development, feedback obtained through the mechanisms described in 4a2.2.1 informed how we specified the measure to align with immunization guidelines from the Advisory Committee on Immunization Practices.

Improvement

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

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

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

This is a new measure; therefore, we do not yet have data on improvement over time. Adoption of this measure has the potential to improve the quality of prenatal care and prenatal immunization rates.

4b2. Unintended Consequences

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

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

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

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

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

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0039 : Flu Vaccinations for Adults Ages 18 and Older

0041 : Preventive Care and Screening: Influenza Immunization

0431 : INFLUENZA VACCINATION COVERAGE AMONG HEALTHCARE PERSONNEL

0680 : Percent of Residents Who Were Assessed and Appropriately Given the Seasonal Influenza Vaccine (Short Stay)

0681 : Percent of Residents Assessed and Appropriately Given the Seasonal Influenza Vaccine (Long Stay)

1659 : Influenza Immunization

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

No

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

This measure specifically assesses immunizations administered during prenatal care. Other related measures assess broader populations and older adults, and do not provide information about the quality of care provided to pregnant women.

5b. Competing Measures

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

OR

Multiple measures are justified.

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

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

Not applicable.

Appendix

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

No appendix Attachment:

Contact Information

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

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

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

Co.4 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

PREGNANCY HEALTH MEASUREMENT ADVISORY PANEL

Alison Chi, American Immunization Registry Association

Annie Fedorowicz, Minnesota Department of Health

Nicole Garro, March of Dimes

Howard Minkoff, Maimonides Medical Center

Renee Miskimmin, Meridian Health Plan

Sarah Royce, California Department of Health

Catherine Ruhl, Women's Health Programs Association of Women's Health

Carol Sakala, National Partnership for Women and Families

Kimberly Sherman, US Department of Health and Human Services

TECHNICAL MEASUREMENT ADVISORY PANEL?

Andy Amster, MSPH, Kaiser Permanente?

Sarah Bezeredi, MBA, MSHL, UnitedHealth Group

Jennifer Brudnicki, MBA, Inovalon Inc.

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

Mike Farina, MBA, R.Ph, Capital District Physicians' Health Plan

Marissa Finn, MBA, CIGNA?

Scott Fox, MS, Med, FAMIA, The MITRE Corporation?

Carlos Hernandez,?CenCal?Health?

Harmon Jordan, ScD, Westat??

Gigi Raney, LCSW, Center for Medicaid and CHIP Services

Lynne?Rothney-Kozlak, MPH,?Rothney-Kozlak Consulting, LLC?

Laurie Spoll, Aetna

?

COMMITTEE ON PERFORMANCE MEASUREMENT??

Andrew Baskin, MD, Aetna??

Elizabeth Drye, MD, SM, Yale School of Medicine

Andrea Gelzer, MD, MS, FACP, AmeriHealth Caritas

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

David Grossman, MD, MPH, Washington Permanente Medical Group

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

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

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

Nancy Lane, PhD, Independent Consultant

Bernadette Loftus, MD, Freelance

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

Amanda Parsons, MD, MBA, Metroplus

Wayne Rawlins, MD, MBA, ConnectiCare

Misty Roberts, MSN, RN, CPHQ, PMP, Humana

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

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

JoAnn Volk, MA, Georgetown University

The NCQA Pregnancy Health Measurement Advisory Panel advised NCQA during measure development. They evaluated the way staff specified the measure, reviewed field test results, and assessed NCQA's overall desirable attributes of Relevance, Scientific Soundness, and Feasibility. The advisory panel consisted of a balanced group of experts, including representatives from pediatric care. In addition to this advisory panel, we vetted the measure with a host of other stakeholders, as is our process. Thus, our measures are the result of consensus from a broad and diverse group of stakeholders.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2018

Ad.3 Month and Year of most recent revision:

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

Ad.5 When is the next scheduled review/update for this measure? 12, 2020

Ad.7 Disclaimers: These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. THE MEASURSE AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

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