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: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 1516

Corresponding Measures:

De.2. Measure Title: Well-Child Visits in the Third, Fourth, Fifth, and Sixth Years of Life

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

De.3. Brief Description of Measure: The percentage of children 3-6 years of age who had one or more well-child visits with a primary care physician during the measurement year.

1b.1. Developer Rationale: This measure encourages health plans to ensure children enrolled in their plans are receiving well-child visits. Well-child visits are an important opportunity to assess a child's health and development; provide immunizations and screenings; and engage parents and caretakers on the health and well-being of their child.

S.4. Numerator Statement: Children who received at least one well-child visit with a PCP during the measurement year.

S.6. Denominator Statement: Children 3-6 years of age during the measurement year.

S.8. Denominator Exclusions: This measure excludes children in hospice.

De.1. Measure Type: Process

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

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 15, 2011 **Most Recent Endorsement Date:** Sep 02, 2015

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

Summary of prior review in [2014]

- Measure is based on updated guideline/expert opinion used in prior submissions.
- The evidence is based on expert consensus and systematic review. No RCTs have been performed.
- As a process measure the evidence should demonstrate that children 3-6 years who have at least one well-child visit each year and receive screenings and anticipatory guidance will experience improved health and an improved quality of life. The evidence to support this measure provided by the developer is derived from the 'Recommendations for Pediatric Preventive Health Care. American Academy of Pediatrics' (AAP) and from the 'Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Pocket Guide'. The AAP/ Bright Futures both recommend at least four well-care visits for children 3-6 years of age.

Changes to evidence from last review

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

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

Updates:

Exception to evidence

N/A

Question for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

Guidance from the Evidence Algorithm

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

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

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

Maintenance measures – increased emphasis on gap and variation

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

- Performance gap data are summarized at the health plan level. In 2016, there was about a 26-percentage point difference between commercial plans in the 10th and 90th percentiles and a 22-percentage point difference for Medicaid plans.

Disparities

- The developer notes that literature documents disparities in well-child visits by race and ethnicity.
  - A study on children born between 2007 and 2009 examined the frequency of well-child visits in infants in relation to their demographics. This study found that children with commercial insurance coverage had a 73.4% compliance rate for receiving their well-child visits versus children with Medicaid who were 45% compliant (Dabney et. al, 2012).
  - Regarding race, Caucasian children were more likely to be compliant as compared to other ethnicities (Caucasian: 68.1%; African American: 46.1%; Asian: 66.3%; Hawaiian/PI: 53.4%; Other: 52.4%). When the authors controlled for insurance status, African American children were still less likely (42%) than Caucasian children (58%) to be compliant (Dabney et. al, 2012).
  - In 2011 to 2012, children and adolescents (from birth to 17 years) from poor (78.5%), low-income (81%), and middle-income (85.9%) households were less likely to have well-child visits than those from high-income (90.3%) households. Black children had lower rates of well-child visits compared with their White counterparts, while Hispanic children had lower rates than White, Black, and other non-Hispanic children (AHRQ, 2014).

Question for the Committee:

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

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

RATIONALE:

- The performance gap in care is significant. Significant gaps exist in care for commercial vs. Medicaid populations..
- Significant disparities between ethnic and socioeconomic status presented based on the literature

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

Evidence:
**The evidence for this very important issue is unfortunately very poor. I would much rather have seen a set of evidence-based procedures needed for children in these age groups that would justify annual visits. Eye and hearing exams and cholesterol measurement, for example, are not needed every year. Once we know the processes that are needed, then one can see if they are best accomplished through "well-child visits" rather than opportunistically or in some other manner. One should also be able to assess whether or not the needed (i.e. evidence based) services are actually delivered rather than simply attendance at a visit
**The evidence clearly supports the process measure and applies directly. The process measure supports the outcome via expert consensus. I am not aware of any new studies or information that changes the evidence base for this measure that has not been cited in the submission.
** Process Measure- Maintenance Measure; No new evidence, still merits value for health plan performance
** The evidence is based on long-standing and well-accepted practice guidelines. These guidelines are not, however, based on RCTs per se, and that has not changed. Therefore, there is no need for a repeat vote.

Performance Gap:
**Yes, there's sufficient variability
**Yes, the current performance data on the measure was provided and demonstrates a moderate gap in care warranting a national performance measure. Yes the data showed significant subgroup variability demonstrating gaps in care by race, ethnicity, and income status.
** Yes there is still evidence that disparities exist among race and ethnicity data.
** There are long-standing and well-established disparities by race, ethnicity, parents' income, and other factors, and between Medicaid and other plans.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing

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

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.

Evaluation of Reliability and Validity:

- The developer conducted empirical score-level reliability and validity testing.
- The developer addressed a former concern and added a new denominator exclusion for patients in hospice. Patients in hospice will not necessarily benefit from preventive care, so it is not appropriate to assess them for receipt of well-child visits.
- Reliability assessment: The developer provides measure score reliability using the beta-binomial model (signal to noise ratio) and 2016 HEDIS data: 405 commercial plans and 243 Medicaid plans comprised the data set
    - The reliability score for commercial plans was 1.00 and the 10-90th percentile range was 0.96-1.00. For Medicaid plans, the statistics were 0.99 and 0.93-1.00, respectively.
    - The developer concludes the measure has high reliability.
    - The developer does not offer a data element level test of reliability for a measure that uses multiple data sources as an option; the measure has an administrative data only option and an administrative data+medical record review option. Reliability of the measure is dependent on correct abstraction of well-child visit data from health records. The developer does not offer an analysis of these data sources relative to the stability of the metric.
- Validity assessment: The developer assessed construct validity against other HEDIS measures and face validity.

- Pearson Correlation Coefficients were calculated for three aspects of weight assessment counseling (BMI percentile, counseling for nutrition, counseling for physical activity); Children Access Primary Care Provider 25 months-6 years; Childhood Immunization Status—All Vaccines; Well-Child Visits in the First 15 Months of Life. Correlation coefficients ranged from 0.4 to 0.8; all correlations significant at $p<0.0001$. The developer concludes the results indicate moderate to strong positive correlation of the hypothesized associations.
- Face validity was assessed by three panels: Child Health Measurement Advisory Panel; Technical Measurement Advisory Panel; Committee on Performance Measurement. The developer states the panels concluded "with good agreement" that the measure is specified to accurately to assess well-child visits in health plans." No quantitative representation of agreement was provided.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- Do you have any concerns that the measure may be calculated in either of two approaches: 1) administrative data only, or 2) administrative data combined with medical record review? Does the Committee wish to discuss the impact, if any, on scores using the different approaches?
- NQF staff is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- NQF staff is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

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

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

Scientific Acceptability

Measure Number: 1516

Measure Title: Well-Child Visits in the Third, Fourth, Fifth, and Sixth Years of Life

Type of measure:

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

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

Data Source:

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

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

☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

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

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

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

   Submission document: "MIF_1516" 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.**

   - None

RELIABILITY: TESTING

Submission document: "MIF_1516" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

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

   ☐ **Yes** ☐ **No**

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

   - The developer provides measure score-level reliability testing using a 2016 HEDIS dataset comprised of 405 commercial plans and 243 Medicaid plans.
   - Developer used the beta-binomial method of testing which the proportion of variation attributable to measurement error for each plan, which represents "noise." The reliability of the measure is represented as the ratio of signal to noise.

   Submission document: Testing attachment, section 2a2.2

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

   - The developer reports that the reliability scores for the *Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life* measure the vast majority of plans met or exceeded the minimally acceptable threshold of 0.7, and the majority of plans exceeded 0.9.

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.**

- The developer provides measure score reliability for *Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life,* with reliability statistics indicating high reliability as defined by Adams (2009) for performing the score-level testing.
- The measure developer does not offer sufficient evidence of data element level reliability, given the different data sources/approach.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

- The developer addressed a former concern and added a new denominator exclusion for patients in hospice. Patients in hospice will not necessarily benefit from preventive care, so it is not appropriate to assess them for receipt of well-child visits.

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

    Submission document: Testing attachment, section 2b4.

- No concerns. Inter-quartile range and t-test to identify meaningful differences was used.
- The results showed a 12.9 percentage point difference between plans in the 25th and 75th percentiles among commercial plans, and a 12.3 percentage point difference among Medicaid plans. For both plan types, the differences between the 25th and 75th percentile were statistically significant.

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.

- Measure use both paper records and/or claims data as data sources.
- As noted earlier, the Committee may wish to discuss the impact on scores using the two options (administrative data only or administrative data+medical records).

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

    Submission document: Testing attachment, section 2b6.

- The developer discusses its audit process and states that "only performances rates that have been reviewed and determined not to be "materially biased" [due to missing data] are reported and used. The developer states its auditing process did not raise any issues regarding missing data for this measure.

16. **Risk Adjustment**

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

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

☐ Yes ☐ No ☐ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d. **Risk adjustment summary:**

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

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

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

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

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

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

For cost/resource use measures ONLY:

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

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

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

VALIDITY: TESTING

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

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

☒ **Face validity**

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

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

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

Submission document: Testing attachment, section 2b2.2

- Face Validity established by multiple NCQA panels (3 panels) who concluded with "good agreement" that the measure is specified accurately to assess well-child visits in health plans.
- Empirical testing centered on score-level assessment of construct validity by using Pearson correlation and comparing the performance on the measure with other measures of childhood care. The following hypothesis testing was conducted:
    - Well-child visits for children 3, 4, 5 and 6 years old will be positively correlated with measures of preventive care for children (Weight Assessment and Counseling: BMI Percentile, Counseling for Nutrition, Counseling for Physical Activity; and Childhood Immunization Status).
    - Well-child visits for children 3, 4, 5 and 6 years old will be positively correlated with other measures of access to care for children (Well-Child Visits in the First 15 Months of Life; Children's Access to Primary Care Practitioners [ages 25 months-6 years]).

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

Submission document: Testing attachment, section 2b2.3

- Pearson Correlation Coefficients were calculated for three aspects of weight assessment counseling (BMI percentile, counseling for nutrition, counseling for physical activity); Children Access Primary Care Provider 25 months-6 years; Childhood Immunization Status—All Vaccines; Well-Child Visits in the First 15 Months of Life. Correlation coefficients ranged from 0.4 to 0.8; all correlations significant at p<0.0001. The developer concludes the results indicate moderate to strong positive correlation of the hypothesized associations. The developer notes that coefficients with absolute values of less than 0.3 are generally considered indicative of weak associations: values of 0.3 or higher denote moderate to strong associations. All childhood measures had values higher than 0.4.
- Multiple NCQA panels (3 panels of experts) concluded with good agreement that the measure is specified accurately to assess well-child visits in health plans. The extent of agreement was not quantitatively reported.

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

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

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

    Submission document: Testing attachment, section 2b1.

    ☐ **Yes**

    ☐ **No**

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

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

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

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

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

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

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

    - Construct validity correlated performance on this measure with other process measures of childhood care were moderate to high.
    - Exclusion of children in hospice is logical. No quantitation of the impact of the exclusion on performance was provided.

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.**

N/A

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

Specifications:
**See comment above (The evidence for this very important issue is unfortunately very poor. I would much rather have seen a set of evidence-based procedures needed for children in these age groups that would justify annual visits. Eye and hearing exams and cholesterol measurement, for example, are not needed every year. Once we know the processes that are needed, then one can see if they are best accomplished through "well-child visits" rather than opportunistically or in some other manner. One should also be able to assess whether or not the needed (ie evidence based) services are actually delivered rather than simply attendance at a visit)
**All data elements are clearly defined and all codes with descriptors are provided. All steps are clear. I have no concerns about the likelihood that this measure can be consistently applied.
**No concerns reliability; The measure does use administrative claims data and some medical review, but as fewer non-EMR systems are in place more emphasis will be placed on non-claims information.
**No concerns.
Reliability Testing:
**None
**No concerns. The beta-binomial method was used with complete data from all NCQA plans, and high reliability scores were obtained.
Validity Testing:
**See above. Not clear that they are getting the key information needed. (The evidence for this very important issue is unfortunately very poor. I would much rather have seen a set of evidence-based procedures needed for children in these age groups that would justify annual visits. Eye and hearing exams and cholesterol measurement, for example, are not needed every year. Once we know the processes that are needed, then one can see if they are best accomplished through "well-child visits" rather than opportunistically or in some other manner. One should also be able to assess whether or not the needed (ie evidence based) services are actually delivered rather than simply attendance at a visit)
**None
**No concerns. Appropriate face-validity and empirical testing methods were used with good results.
Threats to Validity:
**Not a concern
**No concerns regarding validity. Appropriate face-validity and empirical testing methods were used with good results. Statistical tests were performed to demonstrate that the measure easily distinguished care at the 25th and 75th percentiles. The measure may be calculated in either of two approaches: 1) administrative data only, or 2) administrative data combined with medical record review. There is nothing in the submission, however, indicating any testing of the impact, if any, on scores using the different approaches. Since this approach has been the same since the measure was introduced, I don't see any reason to rate the measure down on this score now. It would be useful, however, to request an analysis of this in the next resubmission. Missing data is addressed in the auditing process.
Other Threats to Validity:
**The exclusion of the very small number of children in hospice is consistent with the evidence. 2b3. N/A
**Not applicable for this measure
**The use of this measure has been well demonstrated.

Criterion 3. Feasibility

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

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

- The measure's data are collected through administrative claims, electronic clinical data, and/or paper records.
- The developer anticipates that as electronic health records become more widespread the reliance on paper record review will decrease.

Question for the Committee:

- Does the Committee feel that the measure should be tested with specifications that use electronic data sources only?

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

RATIONALE:

- The measure's data are collected through administrative claims, but the plan may opt to be scored on a claims + medical record option. Because of the reliance on the latter, the preliminary rating for feasibility is moderate.
- The developer anticipates that as electronic health records become more widespread the reliance on paper record review will decrease.

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

** Feasible
** All of the data elements are routinely generated and used during care delivery and are usually available in electronic form. I have no concerns about how the data collection strategy can be put into operational use.
**Long standing measure, feasibility not a concern
** The measure's data are collected through administrative claims, electronic clinical data, with the optional addition of paper records. As electronic health records become more widespread the reliance on paper record review will likely decrease. Although one might question what this change would mean to the measure's scientific acceptability, it would only improve feasibility. So, to the extent that the previous rating was based on the use of paper records, and this has not proven to be a problem, and the reliance on paper records is diminishing anyway, I would rate the feasibility as High.

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

- CMS Medicaid CHIP Child Core Set
    - These are a core set of health quality measures for children enrolled in Medicaid/Children's Health Insurance Program (CHIP) to be reported at the state level. The data collected from these measures will help CMS to better understand the quality of health care that children enrolled in Medicaid/CHIP receive nationally.
- NCQA State of Health Care Annual Report
    - This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2016, HEDIS measures covered 114.2 million commercial health plan beneficiaries and 47.0 million Medicaid beneficiaries
- NCQA Quality Compass
    - This measure is used in Quality Compass which is an indispensable tool used for selecting health plans, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.
- CMS Health Insurance Market Quality Rating System
    - This measure is used in the CMS-developed Quality Reporting Rating System (QRS) set of measures. The QRS measure set consists of measures that address areas of clinical quality management; enrollee experience; and plan efficiency, affordability and management. The measure set includes a subset of NCQA's HEDIS measures and one PQA measure.

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

- NCQA regularly provides technical assistance on measures through its Policy Clarification Support System.
    - Questions received through the Policy Clarification Support system have generally centered around which types of documentation count for plans submitting using the medical record review method.
- During the measure's last major update in 2015, feedback was obtained using a consensus-based process to consider input from multiple stakeholders, including but not limited to entities being measured. Methods to obtain input, include vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System.

Additional Feedback:

During its 2018 review of the current Medicaid Child Core Set, the NQF Medicaid Child Workgroup supported the measure for continued use in the program.

***Question for the Committee**:*

- Can the performance results be used to further the goal of high-quality, efficient healthcare?

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

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

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

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

Improvement results

- The developer reported a slight improvement in performance for commercial plans over the past three years, increasing from 73.71% in 2014 to 75.45% in 2016. Performance was steady over the past three years for Medicaid plans (71.91% in 2014, 72.17% in 2016). These rates suggest opportunity for continued performance improvement.

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

Unexpected findings (positive or negative) during implementation

- The developer did not identify any unintended consequences for this measure during testing or since implementation.

Potential harms

- The developer did not identify any unintended benefits for this measure during testing or since implementation.

Additional Feedback: N/A

***Question for the Committee**:*

- How can the performance results be used to further the goal of high-quality, efficient healthcare?

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

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

Use:
** There seems to be very little progress in improving this metric, so use seems modest at best
** The measure is being publicly reported to health plans, states, CMS, and other public organizations. 4a2. Yes to all three questions.
**This measure is currently is use for many health plan performance programs.
** The usability of this measure has been well demonstrated.
Usability:
** Benefits are not entirely clear, but annual visits are costly in time and money
** The performance results can be used by health plans to report back to providers and provide suggestions for improvement. There are no unintended consequences.

**No concerns for unintended consequences.
** The usability of this measure has been well demonstrated.

Criterion 5: Related and Competing Measures

Related or competing measures
The developer did not identify any measures that are related or competing to this measure.

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

** There are no competing measures. However, measure # 1392, All Child Visits in the First 5 Months seems to be a related measure.
**No competing measures evident

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January 24, 2019
No comments have been received.

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

1516_Evidence_Form-636778984038274615.docx

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

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

No

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1516

Measure Title: Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life

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

Date of Submission: 11/15/2018

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:

☐Patient-reported outcome (PRO):

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

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

☒ Process: Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

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

2015 Submission:

Children 3-6 years >> attend well-child visit >> receive screenings and anticipatory guidance >> improved health 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.)

2018 Submission:

N/A

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

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

2018 Submission:

Not applicable – this is not an outcome measure.

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

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**2018 Submission:** • Recommendations for Preventive Pediatric Health Care. • Bright Futures/American Academy of Pediatrics [AAP] • February 2017 • https://www.aap.org/en-us/Documents/periodicity_schedule.pdf • Bright Futures: Guidelines for Health Supervision of Infants, Children and Adolescents, 4th Edition • Hagan JF, Shaw JS, Duncan P, eds. American Academy of Pediatrics. • February 2017. **2015 Submission:** • 2014 Recommendations for Pediatric Preventive Health Care. • Committee on Practice and Ambulatory Medicine. American Academy of Pediatrics. • March 2014 • http://pediatrics.aappublications.org/content/133/3/568.full.pdf+html • Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents, Third Edition. Pocket Guide. • Hagan JF, Shaw JS, Duncan P, eds. American Academy of Pediatrics. • 2008. • http://brightfutures.aap.org/pdfs/bf3%20pocket%20guide_final.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.	**2018 Submission:** The AAP/Bright Futures guidelines continue to recommend a total of four well-care visits for children ages three to six years of age. https://www.aap.org/en-us/Documents/periodicity_schedule.pdf **2015 Submission:** The AAP/Bright Future guidelines recommends a total of four well-care visits for children ages three to six years of age. They recommend that the well-care visits include, but are not limited to, a medical history, measurements (length/height and weight, BMI, blood pressure), sensory screening (vision and hearing), development surveillance, psychosocial/behavioral assessment, physical exam, immunization, oral health and anticipatory guidance.

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**2018 Submission:** The AAP/Bright Futures did not provide a grading system for the recommendations. To develop the recommendations, the Bright Futures Steering Committee used a similar process as the one used to produce the third edition of the guidelines. 1. Four multidisciplinary expert panels were convened for the age stages of infancy, early childhood, middle childhood, and adolescence. Each panel was co-chaired by a pediatrician content expert and a panel member who represented family members or another health profession. The 39 members of the expert panels were individuals who represented a wide range of disciplines and areas of expertise (e.g., mental health experts, nutritionists, oral health practitioners, family medicine providers, nurse practitioners). 2. An evidence expert advised the Bright Futures Steering Committee and editors. 3. In 2015, the guidelines were posted for public review before publication. External reviewers who represented AAP committees, councils, and sections; professional organizations; institutions; and individuals with expertise and interest in this project provided more than 3,500 comments. **2015 Submission:** The Bright Futures Steering Committee used three different approaches to develop the guidance and recommendations. 1. Multidisciplinary Expert Panels were convened to write recommendations for Bright Futures visit priorities, the physical examination, anticipatory guidance, immunizations, and universal and selective screening topics for each age and stage of development. 2. A Bright Futures Evidence Panel, composed of consultants who are experts in finding and evaluating evidence from clinical studies, was convened to examine studies and systematic evidence reviews and to develop a method of informing readers about the strength of the evidence. 3. Throughout the Guidelines development process, the Project Advisory Committee and Expert Panels consulted with individuals and organizations with expertise and experience in a wide range of topic areas. The entire Guidelines document also underwent public review twice in 2004 and once in 2006. More than 1,000 reviewers, representing national organizations concerned with infant, child and adolescent health. Please see link below for more information on the rationale and evidence for the Bright Futures Recommendations: http://brightfutures.aap.org/pdfs/Guidelines_PDF/13-Rationale_and_Evidence.pdf
Provide all other grades and definitions from the evidence grading system	**2018 Submission:** The AAP/Bright Futures did not provide a grading system for the recommendations. **2015 Submission:** Expert Opinion

Grade assigned to the **recommendation** with definition of the grade	**2018 Submission:** The AAP/Bright Futures did not provide a grading system for the recommendations. **2015 Submission:** No response
Provide all other grades and definitions from the recommendation grading system	**2018 Submission:** The AAP/Bright Futures did not provide a grading system for the recommendations. **2015 Submission:** No response

Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	**2018 Submission:** Quantity: While there have been updates to the evidence supporting individual components of well-child visits (e.g., evidence for specific screenings or interventions), there are no changes in the direction of the evidence that would warrant changes to the measure. https://brightfutures.aap.org/Bright%20Futures%20Documents/BF4_Evidence_Rationale.pdf **2015 Submission:** Quantity: Well child visits for children represent a critical opportunity to conduct a physical examination, administer vaccinations, conduct screenings and provide anticipatory guidance to parents/caregivers. The number and type of study designs included in these recommendations varied and addressed specific aspects of the well-child visit, as few clinical trials assess the visit itself as an intervention that affects health outcomes. Bright Futures included policy statements and published reviews of the American Academy of Pediatrics (AAP) and other national organizations as well as articles from the literature. • One large randomized controlled trial assessed the effect of home visits by paraprofessionals and nurses • No controlled trials of cholesterol screening in children were found • Vision screening from birth to age 5 years is recommended by the U.S. Preventive Services Task Force and based on a comprehensive evidence review. • Hearing screening is promoted by the American Academy of Pediatrics Committee on Practice and Ambulatory Medicine • Blood pressure screening is recommended by the National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents • Vaccinations, which are provided during well-child exams, are recommended by the Advisory Committee on Immunization Practices. In addition, vaccinations, which are provided during well-child exams, are recommended by the Advisory Committee on Immunization Practices. Quality: The guideline developers note that evidence for effectiveness was a core criterion for determining which interventions in child health supervision should be included or excluded. There were several randomized controlled trials considered for specific components of well-child visits, and some components are based on recommendations from national organizations such as the U.S. Preventive Services Task Force and the Advisory Committee on Immunization Practices. However, the guideline developers acknowledge that, for many interventions commonly done in children, few properly constructed studies have been done that are able to link the intervention to a health outcome.

Estimates of benefit and consistency across studies	**2018 Submission:** There are no changes in the direction of the evidence that would warrant changes to the measure. **2015 Submission:** There was consistent positive benefit for conducting universal screening and/or interventions for children during the well-child visit, yet the magnitude of the benefit was undefined in most cases. One benefit was that early screening/intervention confirmed normal development and/or identified developmental risks or disabilities that in some cases led to improved outcomes. The body of evidence reviewed determined that well-child visits are an important opportunity to assess the health and function of child and family. The guideline developers emphasized the benefits of having routine visits with a provider and one that includes appropriate interventions such as obtaining a medical history, administering questionnaires or screening tools, performing a physical examination, and providing anticipatory guidance. Assessing growth and development at most or all well-child visits has the benefit to identify developmental delays and administer appropriate treatment as needed. The guidelines note that desired outcomes of the well-child visit are broader than just detecting disease. Although there is no direct evidence evaluating the effect of a visit on outcomes, evidence does demonstrate the benefit of key elements of the visit.
What harms were identified?	**2018 Submission:** No harms were identified regarding well-child visits. **2015 Submission:** No harms were studied regarding well-child visits, though one large RCT of one versus two exams showed no difference in the use of health care resources between the two groups. However, this study focused on newborn exams. Bright Futures also notes that aspects of repeated examination, including growth monitoring, routine blood pressure measurement and screening for signs of physical and sexual abuse, may result in false positive results that could cause harms. However, the guidelines note that physical examination and developmental evaluation are important opportunities to reassure and educate parents and patients about the range of normal findings, and that performing the examinations could enhance patient (family)-provider communication.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	**2018 Submission:** We did not identify any new studies that draw different conclusions from the systematic review. **2015 Submission:** Not applicable

---

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.

2018 Submission:

There are no changes in the direction of the evidence that would warrant changes to the measure.

2015 Submission:

Early identification of developmental disorders is critical to the well-being of children and their families. It is an integral function of the primary care medical home and an appropriate responsibility of all pediatric health care professionals (AAP, 2006). A well-child visit is a critical opportunity to detect a possible developmental delay or disability and early treatment can lessen the future impact on both the child and family (AAP, 2002). In addition, well-child visits allow physicians to promote behaviors conducive to healthy development, and to give age-appropriate counseling, or anticipatory guidance (Committee on Psychosocial Aspects of Child and Family Health, 2001). Anticipatory guidance given during a well-child visit can change parenting practices (for example, by encouraging the use of time-outs instead of harsher forms of discipline (Sege, 2002), and increase knowledge of injury prevention practices and infant sleep patterns (Child Trends, 2004).

Many children are born with risk factors that predispose them to delayed development and developmental disorders; other children will show delayed or disordered development in early childhood, which if undetected and untreated, can contribute to early school failure and attendant social and emotional problems. Some children will have delayed development attributable to a specific medical condition for which medical treatments may be indicated (AAP, 2006).

Fewer than half of children with developmental delays are identified before starting school (U.S. Department of Education, 2006), by which time significant delays already might have occurred and opportunities for treatment might have been missed. Research shows that early intervention treatment services can greatly improve a child's development. Early intervention services help children from birth through 3 years of age (36 months) learn important skills. Services include therapy to help the child talk, walk, and interact with others (CDC, 2014).

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

2015 Submission:

We conducted a literature review to identify studies relevant to the provision of well-child visits.

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

2015 Submission:

American Academy of Pediatrics. 2006. Identifying Infants and Young Children with Developmental Disorders in the Medical Home: An Algorithm for Developmental Surveillance and Screening. *Pediatrics*, *118*(1), 405 - 420.

American Academy of Pediatrics. 2002. Developmental surveillance and screening of infants and young children. *Pediatrics, 109*(1), 144-145.

Committee on Psychosocial Aspects of Child and Family Health. 2001. The new morbidity revisited: A renewed commitment of the psychosocial aspects of pediatric care. *Pediatrics, 108*(5), 1227-1230.

Regalado, M. & Halfon, N. 2001. Primary care services promoting optimal child development from birth to age three years: Review of the literature. *Archives of Pediatrics and Adolescent Medicine, 155*(12), 1311-1322.

Sege, R. D., Perry, C., Stigol, L. et al. 2002. Short-term effectiveness of anticipatory guidance to reduce early childhood risks for subsequent violence. *Archives of Pediatrics and Adolescent Medicine, 156*(1), 62-66.

Child Trends and Center for Child Health Research. 2004. "Early childhood development in social context." http://www.childtrends.org/?indicators=well-child-visits#_edn1 (Accessed Dec 13, 2014).

U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS), Part C Child Count, 1997–2006.

Centers for Disease Control and Prevention. 2014. "Developmental Monitoring and Screening." http://www.cdc.gov/ncbddd/childdevelopment/screening.html#references (Accessed Dec 13, 2014).

1b. Performance Gap

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

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

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

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

This measure encourages health plans to ensure children enrolled in their plans are receiving well-child visits. Well-child visits are an important opportunity to assess a child's health and development; provide immunizations and screenings; and engage parents and caretakers on the health and well-being of their child.

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

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

The following data demonstrate the variation in the rate of well-child visits for children ages 3, 4, 5 and 6, across Medicaid and commercial health plans. In 2016, there was about a 26-percentage point difference between commercial plans in the 10th and 90th percentiles and a 22-percentage point difference for Medicaid plans. These gaps in performance underscore the opportunity for improvement.

N= Number of plans reporting

Commercial Rate (HMO and PPO Combined)

YEAR	N	MEAN	ST DEV	10TH	25TH	50TH	75TH	90TH	Interquartile Range
2014	401	73.71%	11.51%	59.61%	67.69%	75.09%	81.67%	86.55%	13.986%
2015	408	74.30%	10.84%	61.11%	68.48%	75.96%	81.55%	86.11%	13.07%
2016	405	75.45%	10.38%	61.06%	69.76%	76.75%	82.70%	87.45%	12.94%

Medicaid Rate

YEAR	N	MEAN	ST DEV	10TH	25TH	50TH	75TH	90TH	Interquartile Range
2014	206	71.91%	8.75%	59.62%	65.54%	72.05%	78.46%	83.75%	12.92%
2015	237	71.27%	8.76%	60.64%	64.72%	71.42%	77.57%	82.97%	12.85%
2016	243	72.17%	8.61%	60.71%	66.18%	72.45%	78.51%	82.78%	12.33%

The data references are extracted from HEDIS data collection reflecting the most recent years of measurement for this measure. In 2016, HEDIS measures covered 114.2 million commercial health plan beneficiaries and 47.0 million Medicaid beneficiaries. Below is a description of the denominator for this measure. It includes the number of health plans included in HEDIS data collection and the median eligible population for the measure across health plans.

Commercial

YEAR	N Plans	Median Denominator Size per plan
2014	401	3428
2015	408	3305
2016	405	3333

Medicaid

YEAR	N Plans	Median Denominator Size per plan
2014	206	411
2015	237	411
2016	243	411

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

N/A

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

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

Based on extensive work by NCQA to understand how to promote culturally and linguistically appropriate services among plans and providers, we have many examples of how health plans have used HEDIS measures to design quality improvement programs to decrease disparities in care.

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

A study on children born between 2007 and 2009 examined the frequency of well-child visits in infants in relation to their demographics. This study found that children with commercial insurance coverage had a 73.4% compliance rate for receiving their well-child visits versus children with Medicaid who were 45% compliant. Regarding race, Caucasian children were more likely to be compliant as compared to other ethnicities (Caucasian: 68.1%; African American: 46.1%; Asian: 66.3%; Hawaiian/PI: 53.4%; Other: 52.4%). When the authors controlled for insurance status, African American children were still less likely (42%) than Caucasian children (58%) to be compliant (Dabney et al, 2012).

In 2011 to 2012, children and adolescents (from birth to 17 years) from poor (78.5%), low-income (81%), and middle-income (85.9%) households were less likely to have well-child visits than those from high-income (90.3%) households. Black children had lower rates of well-child visits compared with their White counterparts, while Hispanic children had lower rates than White, Black, and other non-Hispanic children (AHRQ, 2014).

Agency for Healthcare Research and Quality (AHRQ). 2014. "National Healthcare Disparities Report, 2013." http://www.ahrq.gov/research/findings/nhqrdr/nhdr13/chap3.html (Accessed Nov 21, 2014).

Dabney, K., Oceanic, P., Fitzgerald, D., Grant, K. and Holmes, L., 2012. 379 Health Disparities in Well Child Visit in a Comprehensive Pediatric Care Center in United States: Does Insurance Matter?. Archives of Disease in Childhood, 97(Suppl 2), pp.A111-A111.

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):*

Primary Prevention

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

Children

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

N/A

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** 1516_W34_Value_Sets.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

We introduced a new denominator exclusion for patients in hospice. Patients in hospice will not necessarily benefit from preventive care, so it is not appropriate to assess them for receipt of well-child visits.

We have also clarified what documentation counts towards the numerator for plans reporting using medical record review.

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).

Children who received at least one well-child visit with a PCP during the measurement year.

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

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

Health plans can choose to report this measure using 1) administrative claims only or 2) administrative claims supplemented with medical record review (hybrid method). Both methods are described below.

ADMINISTRATIVE:

Patients with at least one well-child visit (Well-Care Value Set) with a PCP during the measurement year.

The well-child visit must occur with a PCP, but the PCP does not have to be the practitioner assigned to the child.

MEDICAL RECORD:

Patients with at least one well-child visit with a PCP during the measurement years. The PCP does not have to be the practitioner assigned to the child.

Documentation from the medical record must include a note indicating a visit with a PCP, the date when the well-child visit occurred and evidence of all of the following:

- A health history.
- A physical developmental history.
- A mental developmental history.
- A physical exam.

- Health education/anticipatory guidance.

Do not include services rendered during an inpatient or ED visit.

Preventive services may be rendered on visits other than well-child visits. Well-child preventive services count toward the measure, regardless of the primary intent of the visit, but services that are specific to an acute or chronic condition do not count toward the measure.

Visits to school-based clinics with practitioners whom the organization would consider PCPs may be counted if documentation of a well-child exam is available in the medical record or administrative system in the time frame specified by the measure. The PCP does not have to be assigned to the member.

The organization may count services that occur over multiple visits, as long as all services occur in the time frame specified by the measure.

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

Children 3-6 years of age during the measurement year.

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

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

Children 3-6 years of age as of December 31 of the measurement year.

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

This measure excludes children in hospice.

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.)*

This measure excludes children who use hospice services (Hospice Value Set) any time during the measurement year.

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

N/A

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

Refer to items S.9 for additional denominator details and attached value sets for codes.

Step 1. Determine the eligible population. To do so, identify children who are 3-6 years of age by December 31 of the measurement year.

Step 2. Search for an exclusion in the patient's history. Exclude these patients from the eligible population.

Step 3. Determine the numerator. To do so, identify patients with at least one well-child visit during the measurement year.

Step 3. Calculate the rate.

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

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

This measure can be reported using Administrative method only or Administrative method supplemented with Medical Record review. For organizations that choose to conduct a Medical Record review, a random sample with a minimum sample size of 411 is used. NCQA provides a Random Number table that organizations use to assist with drawing a random sample.

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

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

N/A

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

If other, please describe in S.18.

Claims, Electronic Health Data, Paper Medical Records

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

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

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

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

No data collection instrument provided

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

Health Plan

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

Outpatient Services

If other:

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

N/A

2. Validity – See attached Measure Testing Submission Form

1516_Testing_Form-636700124002745930.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** 1516
Measure Title: Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life
Date of Submission: 8/15/2018

Type of Measure:

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

Instructions

- 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.
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For 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) 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.

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 evidence and 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.

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 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). N/A

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

2018 Submission: 2016

2015 Submission: 2014

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

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

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

2018 Submission:

Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life is a Healthcare Effectiveness Data and Information Set (HEDIS) health-plan level measure that assesses whether children ages 3, 4, 5 and 6 covered by commercial and Medicaid plans received well-child visits per recommendations. We tested the measure in health plans, which is appropriate for the accountable entity for which the measure is specified.

MEASURE SCORE RELIABILITY TESTING and CONSTRUCT VALIDITY TESTING

The measure score reliability and construct validity testing were calculated from HEDIS data that included 405 commercial plans and 243 Medicaid plans. These totals include all Medicaid and commercial health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size.

SYSTEMATIC EVALUATION OF FACE VALIDITY

The measure was tested for face validity with three panels of experts:

The Child Health Measurement Advisory Panel includes 13 experts in pediatric care, including representation by consumers, health plans, health care providers, academia and policymakers.

The Technical Measurement Advisory Panel includes 12 members, including representation by health plans, methodologists, clinicians and HEDIS auditors.

NCQA's Committee on Performance Measurement (CPM) is an independent, multi-stakeholder panel that oversees measures used in NCQA programs. The panel includes representation by purchasers, consumers, health plans, health care providers, and policy makers. This panel is made up of 21 members and reports to the NCQA Board of Directors. The panel is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

See Additional Information: Ad.1. Workgroup/Expert Panel Involved in Measure Development for names and affiliations of all expert panels involved with this measure.

2015 Submission:

MEASURE SCORE RELIABILITY TESTING

Measure score reliability was calculated from HEDIS data that included all plans submitting data to NCQA for HEDIS: 410 commercial plans and 212 Medicaid plans. The plans were geographically diverse and varied in size.

CONSTRUCT VALIDITY TESTING

Measure score reliability was calculated from HEDIS data that included all plans submitting data to NCQA for HEDIS: 410 commercial plans and 212 Medicaid plans. The plans were geographically diverse and varied in size.

SYSTEMATIC EVALUATION OF FACE VALIDITY

This measure was tested for face validity with three panels of experts:

- The Child Health Measurement Advisory Panel includes 13 experts in pediatric care, including representation by consumers, health plans, health care providers, academia and policymakers.
- The Technical Measurement Advisory Panel includes 12 members, including representation by health plans, methodologists, clinicians and HEDIS auditors.
- NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the HEDIS measurement set and includes representation by purchasers, consumers, health plans, health care providers and policy makers. The CPM is composed of 16 members, is organized and managed by NCQA and reports to the NCQA Board of Directors. The CPM advises NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

See Additional Information: Ad.1. Workgroup/Expert Panel Involved in Measure Development for names and affiliations of all expert panels involved with this measure.

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*)

2018 Submission:

MEASURE SCORE RELIABILITY AND CONSTRUCT VALIDITY TESTING

In 2016, HEDIS measures covered 114.2 million commercial health plan beneficiaries and 47.0 million Medicaid beneficiaries. Data are summarized at the health plan level and stratified by product line (i.e. commercial, Medicaid). Table 1 below describes the number of health plans that reported the measure and the median eligible population for the measure across health plans.

Table 1. Commercial and Medicaid Plans Reporting *Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life*, 2016

Plan Type	Number of Plans	Median Number of Eligible Members Per Plan
Commercial	405	3,333
Medicaid HMO	243	411

2015 Submission:

Patient sample for measure score reliability and validity testing

2014 Data are stratified by product line (i.e. commercial, Medicaid). Below is a description of the sample. It includes the number of health plans included in HEDIS data collection and the median eligible population for the measure across health plans.

Product Type	Number of Plans	Median number of eligible members per plan
Commercial HMO & PPO Combined	410	3,798
Medicaid HMO	212	8,678

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.

2018 Submission:

Reliability and empiric validity of the measure score was tested using the entire HEDIS data submission (described above). For the systematic assessment of face validity, we have described the composition of the expert panels which assessed face validity in Section 1.5 above.

2015 Submission:

The same data were used for reliability and construct validity as described above. In addition, validity was demonstrated through a systematic assessment of face validity. Per NQF instructions, we described the composition of the expert panels that assessed face validity in the data sample questions above.

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

2018 Submission:

No social risk factors were analyzed. However, this measure is specified for commercial and Medicaid health plans, which is a proxy for income and other socioeconomic factors.

2a2. RELIABILITY TESTING

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

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

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

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

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

2018 Submission:

RELIABILITY TESTING OF PERFORMANCE MEASURE SCORE

We utilized the beta-binomial model (Adams 2009) to assess how well one can confidently distinguish the performance of one accountable entity from another. Conceptually, the beta-binomial model is the ratio of signal to noise. The signal is the proportion of the variability in measured performance that can be explained by real differences in performance. The beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS measures. Reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (i.e., noise), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities).

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

2015 Submission:

METHODS FOR BETA-BINOMIAL RELIABILITY TESTING

The beta-binomial method (Adams, 2009) measures the proportion of total variation attributable to a health plan, which represents the "signal". The beta-binomial model also estimates the proportion of variation attributable to measurement error for each plan, which represents "noise". The reliability of the measure is represented as the ratio of signal to noise.

- A score of 0 indicates none of the variation (signal) is attributable to the plan
- A score of 1.0 indicates all of the variation (signal) is attributable to the plan
- A score of 0.7 or higher indicates adequate reliability to distinguish performance between two plans

PLAN-LEVEL RELIABILITY

The underlying formulas for the beta-binomial reliability can be adapted to construct a plan-specific estimate of reliability by substituting variation in the individual plan's variation for the average plan's variation. Thus, the reliability for some plans may be more or less than the overall reliability across plans.

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

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

2018 Submission:

MEASURE-SCORE RELIABILITY

The reliability for the *Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life* measure per the beta-binomial model is shown in Table 2 for the 405 commercial and 243 Medicaid plans.

Table 2. Beta-Binomial Statistic for Commercial and Medicaid Measure Rate

	Commercial			Medicaid		
	Median	Overall Reliability	10-90th Percentile	Median	Overall Reliability	10-90th Percentile
Well-Child Visits in the 3rd, 4th, 5th and 6th Years of Life	0.99	1.00	0.96 – 1.00	0.93	0.99	0.92 – 1.00

2015 Submission:

The reliability for the Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life measure was estimated at the following for 410 commercial and 212 Medicaid plans:

MEASURE-SCORE RELIABILITY

The reliability for this measure as per the beta binomial model is demonstrated in the table below:

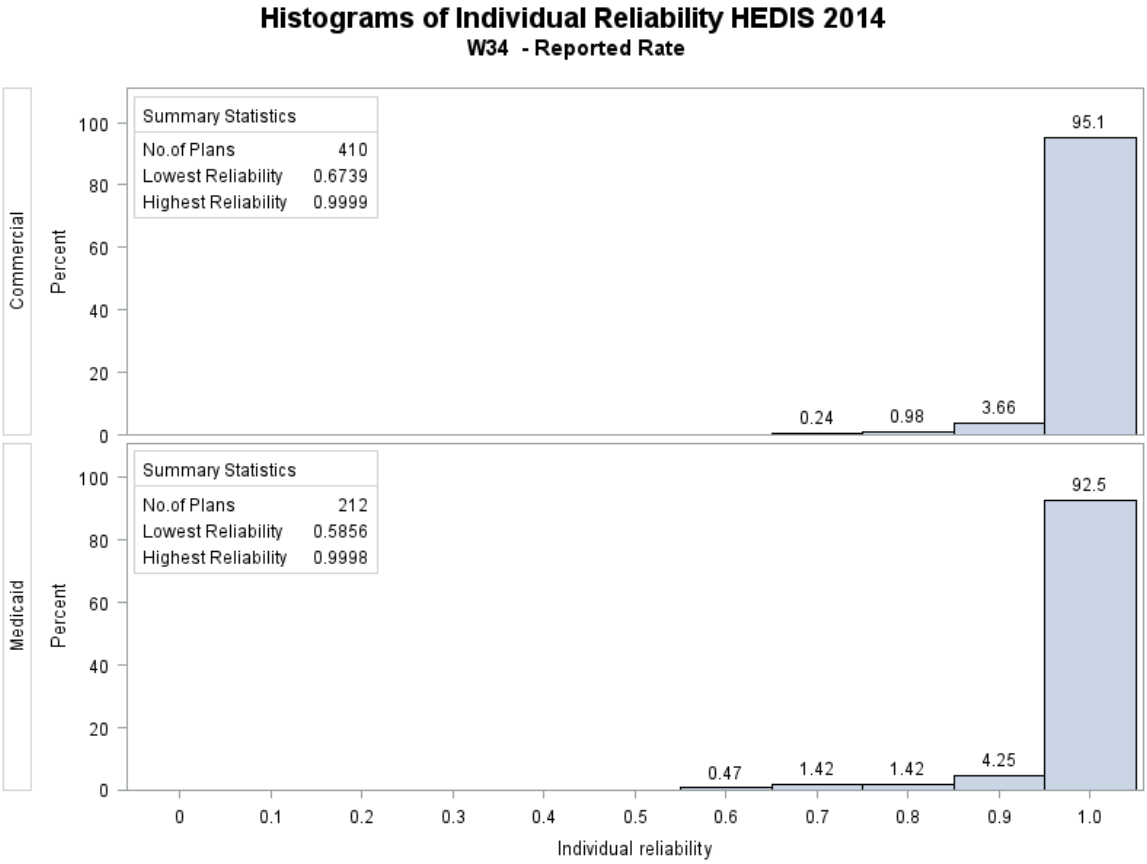
	Overall Reliability
Commercial HMO & PPO Combined	1.00
Medicaid HMO	1.00

PLAN-LEVEL RELIABILITY

This table summarizes the variability of individual plan reliability. The reliability among the 10th percentile plans was above 0.7, indicating high reliability for the majority of plans.

	Average Individual Reliability	Median	10th percentile, 90th percentile
Commercial HMO & PPO Combined	.99	1	.97, 1
Medicaid HMO	.98	1	.97, 1

The histogram shows the distribution of those individual reliability values for each product line.



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?*)

2018 Submission:

Reliability scores can vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (signal). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities.

The testing suggests this measure has good reliability. The 10-90th percentile distribution of health plan level-reliability show the vast majority of health plans met or exceeded the threshold of 0.7, and the majority of plans exceeded 0.9.

2015 Submission:

Results indicate the measure has a strong signal to noise ratio, thus having sufficient signal strength to discriminate performance between accountable entities. Our results suggest the measure is highly reliable. For the well child measures, the vast majority of plans met or exceeded the minimally accepted threshold of 0.7, and the majority of plans exceeded 0.9.

_____

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

2018 Submission:

METHOD OF TESTING CONSTRUCT VALIDITY

We tested construct validity for this measure by comparing the measure with other measures of childhood care. We hypothesized that organizations that perform well on this measure should perform well on other childhood care measures. To test these correlations, we used a Pearson correlation test. This test estimates the strength of the linear association between two variables; the magnitude of correlation ranges from -1 to +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable are associated with decreasing values of the second variable.

For this measure, we hypothesized the following:

1. Well-child visits for children 3, 4, 5 and 6 years old will be positively correlated with measures of preventive care for children (Weight Assessment and Counseling: BMI Percentile, Counseling for Nutrition, Counseling for Physical Activity; and Childhood Immunization Status).
2. Well-child visits for children 3, 4, 5 and 6 years old will be positively correlated with other measures of access to care for children (Well-Child Visits in the First 15 Months of Life; Children's Access to Primary Care Practitioners [ages 25 months-6 years]).

METHOD OF ASSESSING FACE VALIDITY

NCQA develops measures using a standardized process that incorporates the assessment of face validity throughout the process using multi-stakeholder panels. This process is 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.

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. 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. The Board of Directors approved the measure.

2015 Submission:

Method of Assessing Face Validity: NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical expert 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. Refer to What Makes a Measure "Desirable"? The work-up is vetted by NCQA's Measurement Advisory Panels (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. The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

STEP 3: Public Comment is a 30-day period that allows interested parties to offer feedback to NCQA about new measures or changes to existing measures. NCQA MAPs and technical panels consider all comments and advise NCQA staff on recommendations brought to the CPM. The CPM reviews all comments before making a final decision about measures. New measures and changes to existing measures approved by the CPM are included in the next HEDIS year.

STEP 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported nor included in NCQA's State of Health Care Quality, Quality Compass or in accreditation scoring. First-year distinction guarantees that a measure can be effectively collected, reported and audited before it is used for public accountability or accreditation. This is not testing—the measure was already tested as part of its development—rather, it ensures that there are no unforeseen problems during real-world implementation. After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

STEP 5: Public reporting is based on the first-year measure evaluation results. If the measure is approved, it will be publicly reported and may be used for scoring in accreditation.

Step 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reviewed for reevaluation at least every three years. NCQA staff continually monitors the performance of publicly reported measures. Statistical analysis, audit result review and user comments through NCQA's Policy Clarification Support portal contribute to measure refinement during re-evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, NCQA prioritizes measures for re-evaluation and selected measures are researched for changes in clinical guidelines or in the health care delivery systems, and results from previous years are analyzed. Measure work-ups are updated, and the appropriate MAPs review the work-ups and data. If necessary, the measure specifications may be updated or the measure may be recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume 2.

Method of testing construct validity

We tested for construct validity by exploring whether the measure was correlated with measures of quality hypothesized to be related. The Pearson correlation test is used to examine the association between the measures; the test estimates the strength of the linear association between two continuous variables and the magnitude of correlation ranges from -1 and +1, inclusive. A value of 1 indicates a perfect linear dependence in which increasing values on one variable are 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 are 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 the Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life measure, we assessed correlation with the Children and Adolescents' Access to Primary Care Practitioners measure (commercial and Medicaid plans). Our hypothesis was that these two measures would be positively correlated for children greater than 2 years up to 6 years who had the recommended number of preventive care visits (4 visits total) with a PCP. We would expect plans that perform highly on Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life would also perform highly on Children and Adolescents' Access to Primary Care Practitioners for children greater than 2 years up to 6 years.

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

2018 Submission:

CONSTRUCT VALIDITY TESTING: Results are described in Table 3a and 3b below.

Table 3a. Pearson Correlations Among Child Measures in Commercial Plans – 2016

	Pearson Correlation Coefficients					
	Weight Assessment and Counseling			Childhood Immunization Status – All Vaccines	Well-Child Visits in the First 15 Months of Life	Children Access Primary Care Provider – 25 Months – 6 Years
	BMI Percentile	Counseling for Nutrition	Counseling for Physical Activity			
Well-Child Visits in the 3rd, 4th, 5th and 6th Years of Life	0.5	0.5	0.5	0.5	0.7	0.8

Note: All correlations are significant at p<0.0001

Table 3b. Pearson Correlations Among Child Measures in Medicaid Plans – 2016

	Pearson Correlation Coefficients					
	Weight Assessment and Counseling			Childhood Immunization Status – All Vaccines	Well-Child Visits in the First 15 Months of Life	Children Access Primary Care Provider – 25 months – 6 Years
	BMI Percentile	Counseling for Nutrition	Counseling for Physical Activity			
Well-Child Visits in the 3rd, 4th, 5th and 6th Years of Life	0.5	0.6	0.6	0.4	0.5	0.7

Note: All correlations are significant at p<0.0001

2015 Submission:

Construct Validity: Pearson Correlation Coefficient results are shown in Table 1-2.

Table 1. Correlation between Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life and Children and Adolescents' Access to Primary Care Practitioners, Commercial 2014

		Pearson Correlation Coefficients				
		Children and Adolescents' Access to Primary Care Practitioners				Well-Child Visits in the 3rd, 4th, 5th, 6th Years of Life
Commercial		12-24 mos	25 mos-6 yrs	7-11 yrs	12-19 yrs	
Well-Child Visits in the 3rd, 4th, 5th, 6th Years of Life		0.41	0.80	0.79	0.80	1
Children and Adolescents' Access to Primary Care Practitioners	12-24 mo	1				0.41
	25 mo - 6 yrs		1			0.80
	7 – 11 yrs			1		0.79
	12 – 19 yrs				1	0.80

Note: All correlations are significant at p<0.05

Table 2. Correlation between Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life and Children and Adolescents' Access to Primary Care Practitioners, Medicaid 2014

		Pearson Correlation Coefficients				
		Children and Adolescents' Access to Primary Care Practitioners				Well-Child Visits in the 3rd, 4th, 5th, 6th Years of Life
Medicaid		12-24 mos	25 mos-6 yrs	7-11 yrs	12-19 yrs	
Well-Child Visits in the 3rd, 4th, 5th, 6th Years of Life		0.36	0.65	0.48	0.38	1
Children and Adolescents' Access to Primary Care Practitioners	12-24 mo	1				0.36
	25 mo - 6 yrs		1			0.65
	7 – 11 yrs			1		0.48
	12 – 19 yrs				1	0.38

Note: All correlations are 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?*)

2018 Submission:

FACE VALIDITY

Multiple NCQA panels concluded with good agreement that the measure is specified accurately to assess well-child visits in health plans. This measure meets the test for face validity.

CONSTRUCT VALIDTY

Coefficients with absolute values of less than 0.3 are generally considered indicative of weak associations: values of 0.3 or higher denote moderate to strong associations. As shown in Tables 3a and 3b, *Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life* is moderately to highly correlated to the four measures that assess childhood care in both commercial and Medicaid plans. All correlations were significant (p< 0.0001).

2015 Submission:

FACE VALIDITY

Multiple NCQA panels concluded with good agreement that the measures are specified accurately as possible to assess well-child visits in health plans. This measure meets the test for face validity.

CORRELATIONS

As hypothesized, the results in Table 1 show that the Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life measure is strongly positively correlated to the Children and Adolescents' Access to Primary Care Practitioners for children greater than 2 years up to 6 years at 0.80 for commercial plans. The results in Table 2 show that the Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life measure is also strongly positively correlated to the Children and Adolescents' Access to Primary Care Practitioners for children greater than 2 years up to 6 years at 0.65 for Medicaid plans. All correlations were significant (p< 0.05).

_____

2b2. EXCLUSIONS ANALYSIS

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

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

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*)

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*)

_____

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

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

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

☒ No risk adjustment or stratification

☐ Statistical risk model with risk factors

☐ Stratification by risk categories

☐ Other,

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

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

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

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

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

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

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

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

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

If stratified, skip to 2b3.9

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

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

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

2b3.9. Results of Risk Stratification Analysis:

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

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

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

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

2018 Submission:

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25th and 75th percentile on a measure. To determine if this difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 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 measured entities. However, the method can be used for comparison of any two measured entities.

2015 Submission:

To demonstrate meaningful differences in performance, NCQA calculates an interquartile range (IQR) for each indicator. The IQR is a measure of the dispersion of performance and is the difference between the 25th and 75th percentiles on a measure. To determine if the difference is statistically significant, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentiles. This 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' performances are significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans. We used these two plans as examples of 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*)

2018 Submission:

Table 4. *Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life*, Variation in Performance Across Health Plans, 2016

	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Commercial	3,615	75.4	10.4	61.1	69.8	76.7	82.7	87.4	12.9	0
Medicaid HMO	10,010	72.2	8.6	60.7	66.2	72.4	78.5	82.7	12.3	<.0001

EP: Eligible Population, the average denominator size across plans submitting to HEDIS

IQR: Interquartile Range

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

2015 Submission:

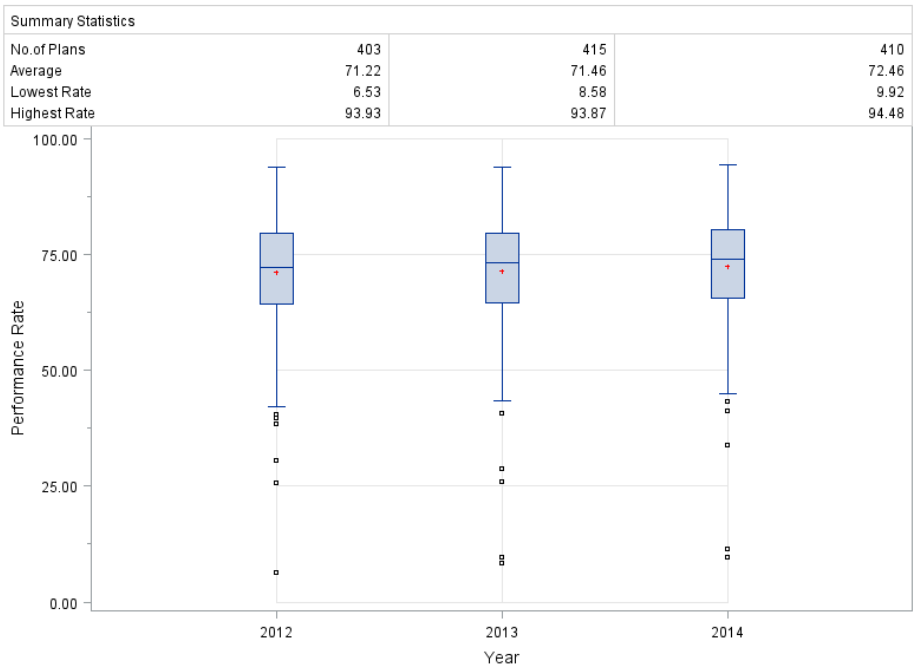
2014 Variation in Performance across Health Plans

	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Commercial HMO & PPO Combined	10,091	72.5	11.6	57.7	65.8	74.0	80.5	86.2	14.7	p<0.05
Medicaid HMO	16,067	71.5	8.7	60.2	66.0	71.8	77.2	82.7	11.2	p<0.05

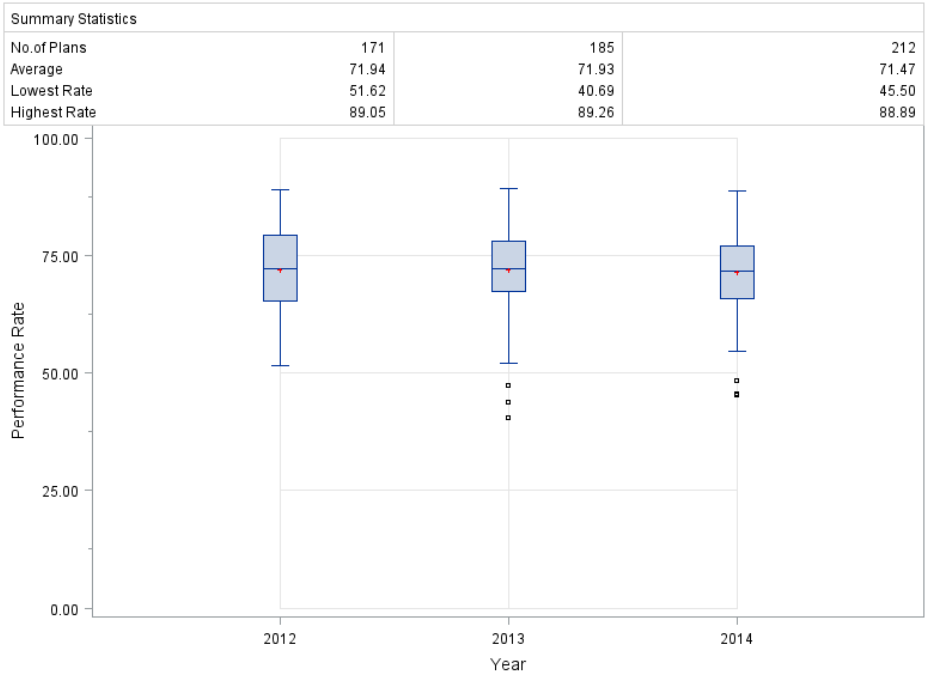
EP: Eligible Population, the average denominator size across plans submitting to HEDIS

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

Boxplot Graph for Commercial W34 Reported Rate from HEDIS 2012-2014



Boxplot Graph for Medicaid W34 Reported Rate from HEDIS 2012-2014



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?*)

2018 Submission:

The results above indicate there is a 12.9 percentage point difference between plans in the 25th and 75th percentiles among commercial plans, and a 12.3 percentage point difference among Medicaid plans. For both plan types, the differences between the 25th and 75th percentile were statistically significant. This gap between the higher and lower performing plans represents an average difference of 466 children (commercial) and 1,231 children (Medicaid) receiving well-child visits at age 3, 4, 5 and 6.

2015 Submission:

Average performance for the measure was 73% for commercial plans (with 10th percentile rate at 58%) and 72% for Medicaid plans (with 10th percentile rate at 60%). The results show an 11-15% gap in performance between the 25th and 75th percentile-performing plans for the measure, which was statistically significant for all product lines and rates. Commercial plans had the largest performance gap with a 14.7 percentage point gap between the 25th and 75th percentiles. All results suggest opportunities for improvement.

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

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

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

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

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and 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*)

2018 Submission:

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

The HEDIS Compliance Audit addresses the following functions:

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

2015 Submission:

This measure is collected with a complete sample; there are no missing data on this measure

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*)

2018 Submission:

HEDIS addresses missing data in a structured way through its audit process. HEDIS measures apply to enrolled members in a health plan, and NCQA-certified auditors use standard audit methodologies to assess whether data sources are missing data. If a data source is found to be missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used. Once measures are added to HEDIS, NCQA conducts a first-year analysis to assess the measure's feasibility once 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.

2015 Submission:

This measure is collected with a complete sample; there are no missing data on this measure

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*)

2018 Submission:

NCQA's auditing procedures did not raise any issues regarding missing data for this measure.

2015 Submission:

This measure is collected with a complete sample; there are no missing data on this measure

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**.

Some data elements are in defined fields in electronic sources

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

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

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

Attachment:

3c. Data Collection Strategy

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

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

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

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

The HEDIS Compliance Audit addresses the following functions:

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

In addition to the HEDIS audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system, NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the 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 Medicaid Child Core Set https://www.medicaid.gov/medicaid/quality-of-care/downloads/medicaid-and-chip-child-core-set-manual.pdf The State of Health Care Quality 2014 http://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality Regulatory and Accreditation Programs Health Insurance Marketplace: Quality Rating System (2015) http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/2015-QRS-Measure-Technical-Specifications.pdf Quality Improvement (external benchmarking to organizations) Quality Compass http://www.ncqa.org/hedis-quality-measurement/quality-measurement-products/quality-compass Quality Rating System Measure Technical Specifications (2018) https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/Revised_QRS-2018-Measure-Tech-Specs_20170929_508.pdf

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

CMS MEDICAID/CHIP CHILD CORE SET: These are a core set of health quality measures for children enrolled in Medicaid/Children's Health Insurance Program (CHIP) to be reported at the state level. The data collected from these measures will help CMS to better understand the quality of health care that children enrolled in Medicaid/CHIP receive nationally.
NCQA STATE OF HEALTH CARE ANNUAL REPORT: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2016, HEDIS measures covered 114.2 million commercial health plan beneficiaries and 47.0 million Medicaid beneficiaries
NCQA QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting health plans, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.
CMS HEALTH INSURANCE MARKET QUALITY RATING SYSTEM: This measure is used in the CMS-developed Quality Reporting Rating System (QRS) set of measures. The QRS measure set consists of measures that address areas of clinical quality management; enrollee experience; and plan efficiency, affordability and management. The measure set includes a subset of NCQA's HEDIS measures and one PQA measure.

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

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

N/A

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

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

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

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

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

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

Describe how feedback was obtained.

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

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

Questions received through the Policy Clarification Support system have generally centered around which types of documentation count for plans submitting using the medical record review method.

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

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

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

During the measure's last update in 2015, feedback obtained through the mechanisms described in 4a2.2.1 informed how we revised the measure.

Improvement

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

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

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

There has been slight improvement in performance for commercial plans over the past three years, increasing from 73.71% in 2014 to 75.45% in 2016. Performance was steady over the past three years for Medicaid plans (71.91% in 2014, 72.17% in 2016). These rates suggest opportunity for continued performance improvement.

4b2. Unintended Consequences

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

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

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

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

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

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5b. Competing Measures

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

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A

Appendix

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

No appendix **Attachment:**

Contact Information

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

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

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

Co.4 Point of Contact: Kristen, Swift, swift@ncqa.org, 202-955-5174-

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.

CHILD HEALTH MEASUREMENT ADVISORY PANEL

Jeanne Alicandro, MD, IPRO

Lekisha Daniel-Robinson, Centers for Medicare and Medicaid Services

Denise Dougherty, PhD, Agency for Healthcare Research and Quality

Theodore Ganiats, MD, University of California, San Diego

Foster Gesten, MD, New York State Department of Health

Charles Homer, MD, MPH, National Initiative for Children's Healthcare Quality

Elizabeth Siteman, Bridges to Excellence

Mary McIntyre, MD, MPH, Alabama Department of Public Health

Virginia Moyer, MD. MPH, FAAP, American Board of Pediatrics

Kevin Park, MD, Molina Healthcare, Inc.

Edward Schor, MD, The Commonwealth Fund

Xavier Sevilla, MD, FAAP, Whole Child Pediatrics

Janet Sullivan, MD, Hudson Health Plan

COMMITTEE ON PERFORMANCE MEASUREMENT

Bruce Bagley, MD, FAAFP, Independent Consultant

Andrew Baskin, MD, Aetna

Jonathan D. Darer, MD, Siemens Healthineers

Helen Darling, MA, Strategic Advisor on Health Benefits & Health Care

Andrea Gelzer, MD, MS, FACP, AmeriHealth Caritas

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

David Grossman, MD, MPH, Washington Permanente Medical Group

Christine Hunter, MD, (Co-Chair) US Office of Personnel Management

Jeffrey Kelman, MMSc, MD, United States Department of Health and Human Services

Nancy Lane, PhD, Independent Consultant

Bernadette Loftus, MD, The Permanente Medical Group

Adrienne Mims, MD, MPH, Alliant Quality

Amanda Parsons, MD, MBA, Montefiore Health System

Wayne Rawlins, MD, MBA, ConnectiCare

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

Eric C. Schneider, MD, MSc (Co-Chair), The Commonwealth Fund

Marcus Thygeson, MD, MPH, Adaptive Health

JoAnn Volk, MA, Reforms

Lina Walker, PhD, AARP

The NCQA Child 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: 1995

Ad.3 Month and Year of most recent revision: 07, 2018

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

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

Ad.7 Disclaimers: This HEDIS® performance measure is not a clinical guideline and does not establish a standard of medical care and has not been tested for all potential applications.

THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures, without modification, are encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Modifications to, and/or commercial use of, a measure requires the prior written consent of NCQA and is subject to a license at the discretion 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.