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.

Purple text represents the responses from measure developers. Red text denotes developer information that has changed since the last measure evaluation review. Some content in the document is from Measure Developers.

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

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 0108

Corresponding Measures:

De.2. Measure Title: Follow-Up Care for Children Prescribed ADHD Medication (ADD)

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

De.3. Brief Description of Measure: Percentage of children newly prescribed attention-deficit/hyperactivity disorder (ADHD) medication who had at least three follow-up care visits within a 10-month period, one of which is within 30 days of when the first ADHD medication was dispensed.

An Initiation Phase Rate and Continuation and Maintenance Phase Rate are reported.

1b.1. Developer Rationale: Attention-deficit/hyperactivity disorder (ADHD) is a brain disorder marked by an ongoing pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development. Medications can improve function, but proper monitoring is recommended. The intent of this measure is to ensure timely and continuous follow-up visits for children who are newly prescribed ADHD medication. The goal is to encourage monitoring of children for medication effectiveness, occurrence of side effects and adherence.

S.4. Numerator Statement: Among children newly prescribed ADHD medication, those who had timely and continuous follow-up visits.

S.6. Denominator Statement: Children 6-12 years of age newly prescribed ADHD medication.

S.8. Denominator Exclusions: Children who had an acute inpatient encounter for mental health or chemical dependency following the Index Prescription Start Date

Children with a diagnosis of narcolepsy: Many of the medications used to identify patients for the denominator of this measure are also used to treat narcolepsy. Children with narcolepsy who are pulled into the denominator are then removed by the narcolepsy exclusion.

Children using hospice services during the measurement year. Children in hospice may not be able to receive the necessary follow-up care.

De.1. Measure Type: Process

S.17. Data Source: Claims

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Jun 28, 2017

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

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

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

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

Summary of prior review in 2016

- The developer provided a rationale for timely follow-up visits after newly prescribing ADHD medication.
- The developer cited systematic reviews in the form of AAP clinical practice guidelines (Strong recommendation; grade B evidence) and AACAP practice parameters ("minimal standard" recommendation; evidence graded rct and ut) for the treatment of ADHD in children and adolescents.
- In the last review, the committee questioned the evidence supporting the 30-day timeframe and its linkage to improved outcomes, and noted barriers to meeting this requirement; the developer said the clinical guidelines support the 30-day period. The committee agreed that the measure addresses a high priority, as ADHD is one of the most prevalent behavioral health diseases in children.

- The developer states: “Numerous (>100) studies related to the care for patients with ADHD have been published since the publication of this guideline, none of which contradict the need for appropriate follow-up once treatment with medication begins.”

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:

- Developer presents evidence from 2019 update of American Academy of Pediatrics (AAP) ADHD guideline, ADHD: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents
    - Key Action Statement (KAS) 1: The pediatrician or other primary care clinicians (PCC) should initiate an evaluation for ADHD for any child or adolescent age 4 years to the 18th birthday who presents with academic or behavioral problems and symptoms of inattention, hyperactivity, or impulsivity. Grade B: Strong Recommendation
    - KAS 4: ADHD is a chronic condition; therefore, the PCC should manage children and adolescents with ADHD in the same manner that they would children and youth with special health care needs, following the principles of the chronic care model and the medical home. Grade B: Strong Recommendation
    - KAS 5b: For elementary and middle school-aged children (age 6 years to the 12th birthday) with ADHD, the PCC should prescribe FDA-approved medications for ADHD, along with parent training in behavior management (PTBM) and/or behavioral classroom intervention (preferably both PTBM and behavioral classroom interventions). Educational interventions and individualized instructional supports, including school environment, class placement, instructional placement, and behavioral supports, are a necessary part of any treatment plan and often include an IEP or a rehabilitation plan (504 plan). Grade A: Strong Recommendation
- Developer relies on previous submission for remaining evidence.

Questions for the Committee:

- Does the Committee agree that the updated evidence provided by the developer is directionally the same and stronger compared to that for the previous NQF review?
- What is the relationship of this measure to patient outcomes? How strong is the evidence for this relationship?

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3)→QQC presented (Box 4)→Quantity: high; Quality: moderate (Grade B evidence); Consistency: high (Box 5b)→**Moderate** (NQF Measure Evaluation Criteria Sept 2019, Algoithm 1 pg. 15)

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.

- Developer presents data extracted from HEDIS data collection reflecting the most recent years of measurement.
    - Performance data are summarized at the health plan level and summarized by mean, standard deviation, minimum health plan performance, maximum health plan performance and performance at the 10th, 25th, 50th, 75th and 90th percentile.
    - Data are stratified by year and product line (i.e. commercial, Medicare, Medicaid) with total plans between ~150-350 per year reported.
    - For HEDIS 2019 (calendar year 2018), HEDIS measures covered 116 million commercial health plan members and 54 million Medicaid enrollees.
    - Performance means range between 40-55% (SD 8-13%).
    - 2017-2019 Performance Data Initiation Phase
        - Commercial

          YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH
          2017 I 39.5% | 8.2% | 11.3% | 75.6% | 30.5% | 35.0% | 38.6% | 43.6% | 50.0%
          2018 I 40.7% | 8.1% | 22.2% | 75.6% | 31.2% | 35.9% | 41.2% | 44.3% | 51.8%
          2019 I 40.0% I 8.3% I 14.0% I 70.5% I 29.8% I 35.0% I 40.0% I 44.3% I 50.4%
        - Medicaid

          YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH
          2017 I 44.5% | 10.4% | 0.0% | 86.1% | 31.7% | 39.0% | 44.8% | 51.8% | 57.1%
          2018 I 44.6% | 9.1% | 20.4% | 73.4% | 34.3% | 38.2% | 45.0% | 50.8% | 55.9%
          2019 I 44.2% I 9.7% I 22.5% I 92.5% I 33.9% I 37.9% I 43.4% I 49.9% I 56.6%
    - 2017-2019 Performance Data Continuation and maintenance phase
        - Commercial

          YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH
          2017 I 46.1% | 8.6% | 25.7% | 74.0% | 36.4% | 40.3% | 44.7% | 51.2% | 57.1%
          2018 I 47.2% | 8.7% | 22.7% | 76.6% | 33.2% | 41.1% | 45.9% | 52.9% | 59.0%
          2019 I 48.0% I 9.3% I 19.1% I 74.1% I 36.2% I 43.2% I 47.9% I 53.0% I 60.0%
        - Medicaid

          YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH
          2017 I 54.5% | 12.9% | 0.0% | 76.9% | 37.3% | 48.2% | 55.9% | 63.7% | 69.5%
          2018 I 55.0% | 11.8% | 20.4% | 76.7% | 33.1% | 41.7% | 50.0% | 63.7% | 69.1%
          2019 I 54.6% I 12.0% I 23.8% I 100.0% I 39.0% I 46.4% I 55.5% I 62.7% I 71.2%

Disparities

- Developer does not provide disparities data and offers a rationale for why this is challenging for health plan data.
- Developer summarizes the literature related to disparities and ADHD
    - Studies suggest children from minority families experience decreased access to and utilization of health services for ADHD, even after controlling for poverty and health insurance status.
    - Although the prevalence of ADHD in minority children is believed to be equal to or even greater than the prevalence in non-minority children, studies indicate that rates of both diagnosis and treatment of ADHD are much lower among minority children compared to non-minority children. Specifically, children who are black, are raised in primarily non-English speaking households, have limited access to the health care system, and are poorer.

Questions 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

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

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

- While evidence suggests follow up is clearly necessary to deal with dosing, side effects, and effectiveness, the schedule of follow up visits as embodied in this measure, and the type of follow up (in person vs telephonic) do not seem to be evidence based.
- This is a process measure based on reviewing claims data. It does not appear to measure other critical data such as change in dosage, discontinuation, or switching to a different medication.
- Evidence is strong.
- This is a maintenance measure. New data from AAP guidelines is sufficient.
- There is good evidence to support the benefits of regular reliable follow up for psycho social interventions as well as medication interventions. The exact timing of the follow-up has not been adequately studied to show that the particular intervals embedded in this measure are evidenced based.
- Updated evidence is provided; I did not see evidence that specifically supported the time frames in the measure (30 days and 3 follow up visits in 10 months).
- There is evidence to support the measure.
- Ongoing concerns about the 1 month requirement. I'm not finding that specific recommendation in the guidelines, and I'm not seeing any new evidence that supports the 1 month specifier as improving outcomes.
- Allows for one of the follow-up visits during C& M phase to be telephone contact which does not allow for adequate medication safety monitoring for children receiving Schedule II drugs.
- Guidelines continue to support.
- Process measure

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

- There appears to be a longstanding performance gap, and there are clear disparities in care which warrant ongoing reporting and addressing.
- It is unclear from the submission if there is a performance gap addressed in the measure.
- Performance gap is evident.
- Performance means range between 40-55% (SD 8-13%).
- There is adequate proof of a genuine performance gap.
- Data was not supplied; however literature supports a performance gap. Data supplied noted performance could be optimized.
- There does seem to be a gap and some disparities re: income we reported.
- Gap exists, espicially with minorities.

- Yes, and very little change in performance gap/10 years.
- There is still a gap.
- There is a performance gap. No data available on disparities.

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.

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

Evaluators: NQF Behavioral Health and Substance Use Staff

Staff Review

Evaluation Summary:

Reliability

- Developer performed reliability testing using a signal-to-noise analysis with the beta-binomial methodology.
- This is a common approach used for reliability testing at the score level for measures of pass/fail events.
- While the developer has only reported a single summary statistic, developer has met with NQF staff to request the opportunity to share a new approach to include confidence intervals around score level reliability point estimates prior to the Committee meeting.
    - Developer was referred to NQF submission criteria: "For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred." (NQF Measure Evaluation Criteria, pg. 18)
- Developer offered the table below as the results:

Measure Rate	Signal-to-Noise Reliability	
	Commercial	Medicaid
Initiation Phase	0.88	0.98
Continuation and Maintenance Phase	0.74	0.94

Validity

- Developer offered two forms of testing
  - Face validity
  - Construct validity using correlation statistics between the individual rates of the measures, as well as with a measure in a comparable quality domain
- Both approaches are commonly used and appropriate forms of score-level testing
- Face validity results found the measure to be sound
- Statistical results from Pearson's Correlation analysis comparing the measure's two rates:
  - Commercial: 0.78
  - Medicaid: 0.89
- Statistical results from Pearson's Correlation analysis comparing the measure with *Use of First-line Psychosocial Care for Children and Adolescents on Antipsychotics*:
  - Commercial: Initiation: 0.26; Continuation and Maintenance: 0.14
  - Medicaid: Initiation: 0.31; Continuation and Maintenance: 0.30
- The individual rates of the measures were highly correlated. Correlation between the measure rates and the comparator measures were weak to moderate.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The NQF staff identified a number of concerns related to reliability that the developer chose not to respond to, namely the unexplained "overall reliability" term, and the spread of the reliability data. Do you feel that you have sufficient information to consider the reliability of this measure?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The measure demonstrated weak to moderate correlations with an external measure which were especially low with the commercial plans. Is this a concern?

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

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

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

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

- While reliability specs and testing is imperfect, it seems adequate to me.
- Given that this is a review of claims data, it can easily be consistently implemented.
- Elements clearly defined.
- Measure specifications and logic model are clear.
- It's reliable and has been implemented reasonable well.
- I am concerned that the measure does not allow for telehealth visits. I think that not allowing for this really could skew the results.
- No issues.
- Specifications clearly defined
- What led to changing the signal to noise calculations?
- No concerns about consistent implementation.

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

- See above.
- No.
- Concern - lack of response from developer regarding unexplained “overall reliability.”
- For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported.
- Could we ask the developer why the signal to noise reliability has declined for the commercial population?
- I am satisfied with the data presented; I do have concerns about televisits being excluded
- No.
- Remains only at the health plan level.
- I would like NCQA and the committee to discuss the change in methodology measuring signal to noise and the resulting .626 for the continuation and maintenance portion of the measure.
- The measure has good reliability.

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

- No concerns with the result. However, this is only a process measure and does not assess outcomes that matter to patients and their families.
- No concerns.
- Correlation between the measure rates and the comparator measures were weak to moderate.
- No.
- Validity testing passable.
- No.
- Concerns about the validity about the 1 month (vs 6 weeks for example) follow up leading to better outcomes.
- Little scientific evidence to support association of adherence to clinical outcomes.
- No.

- The measure has face and construct validity.

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

- I think the issue of risk adjustment is very important if used to evaluate plan/population performance rates. There are clearly health disparities in care that presumably affect overall performance rates.
- I see no inappropriate exclusions.
- No concerns.
- Exclusion of any patient with an inpatient stay. Is it possible that the inpatient stay resulted from lack of follow up care on an outpatient basis. The measure seems to exclude patients who may have had the most severe negative outcome from lack of follow up care.
- Nothing new.
- As noted not permitting telehealth codes.
- No threats.
- Why just 6-12 yo?
- No.
- No concerns.
- Analysis indicates acceptable results.

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

- I worry about alternative follow up approaches and how sensitive claims data are to capturing follow up when ADHD care is embedded with well child care and other visits not coded as ADHD.
- This measure does not appear to me to assess differences in quality.
- No concerns.
- No Issues.
- Could we ask the developer why they felt that comparison to biopsychosocial interventions for children and adolescents on antipsychotics makes sense?
- No concerns.
- No threats.
- High attrition rate between initial and C& M phase.
- No.
- No concerns.

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.

- Developer states that all data are captured during the routine provision of care and are in defined fields within claims data.
- Developer states that NCQA provides a system to allow "real-time" feedback from measure users through a Policy Clarification Support System that receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds to questions in order to prevent possible errors or inconsistencies in the implementation of the measure.
    - Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications.
    - Measures are re-evaluated on a periodic basis and when there is a significant change in evidence.
    - During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the usability and feasibility of the measure.
- Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer.

Questions for the Committee:

- Does the Committee agree that this claims measure is low burden?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- Generally feasible, but prone to missing care episodes due to coding.
- I see no problems with feasability.
- No concerns.
- Claims data.
- No problem.
- Claims data- so feasible.
- I did not find problems with any of the data elements.
- Feasible.
- No concerns.
- Very feasible.

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used

for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided?4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- I am ok with overall use.
- It appears that this measure has been submitted previously and is currently being disclosed widely.
- Publicly reported, no concerns with use.
- Yes.
- No problems here.
- Data is used for a variety of publically reported venues. NCQA measures are regularly re-evaluated.
- Results have been shared.
- I'm happy to see telemed/phone visits being included.
- Yes, and limitations of this measure are published in peer review papers.
- Publicly reported.
- Uses publicly reported data.

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**

Accountability program details

- Measure developer details a number of accountability applications for this measure:
- Public Reporting
  - Health Plan Ranking
- Payment Program
  - CMS Quality Payment Program
  - CMS Promoting Interoperability Programs
  - Physician Value-Based Payment Modifier (VBM)
- Regulatory and Accreditation Programs
  - NCQA Plan Accreditation

- NCQA Accountable Care Organization Accreditation
- Quality Improvement (external benchmarking to organizations)
  - Quality Compass
  - Annual State of Health Care Quality

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

- Developer states that NCQA measures are evaluated regularly through a "reevaluation" process to seek broad input on measures including input on performance and implementation experience including
  - Vetting of the measure with several multi-stakeholder advisory panels
  - Public comment posting
  - Review of questions submitted to the Policy Clarification Support System
- NCQA uses the information to assess a measure's adherence to the HEDIS Desirable Attributes of Relevance, Scientific Soundness and Feasibility.
- Developer states that: "In general, health plans have not reported significant barriers to implementing this measure, as it uses the administrative data collection method. Questions have generally centered around minor clarification of the specifications, such as confirmation that information in claims meets the measure intent and questions about the supporting guidelines for the measure."

Additional Feedback: None

Questions for the Committee:

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

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

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

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

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

Improvement results

- Developer provided 2017 to 2019 data that shows relatively stable performance and room for improvement across Commercial and Medicaid plan with some evidences of improvement over time in the Commercial analysis, but little in the Medicaid analysis.
- 2017-2019 Performance Data Initiation Phase
  - Commercial
    YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH

2017 I 39.5% | 8.2% | 11.3% | 75.6% | 30.5% | 35.0% | 38.6% | 43.6% | 50.0%
2018 I 40.7% | 8.1% | 22.2% | 75.6% | 31.2% | 35.9% | 41.2% | 44.3% | 51.8%
2019 I 40.0% I 8.3% I 14.0% I 70.5% I 29.8% I 35.0% I 40.0% I 44.3% I 50.4%

- Medicaid

  YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH
  2017 I 44.5% | 10.4% | 0.0% | 86.1% | 31.7% | 39.0% | 44.8% | 51.8% | 57.1%
  2018 I 44.6% | 9.1% | 20.4% | 73.4% | 34.3% | 38.2% | 45.0% | 50.8% | 55.9%
  2019 I 44.2% I 9.7% I 22.5% I 92.5% I 33.9% I 37.9% I 43.4% I 49.9% I 56.6%

- 2017-2019 Performance Data Continuation and maintenance phase
  - Commercial

    YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH
    2017 I 46.1% | 8.6% | 25.7% | 74.0% | 36.4% | 40.3% | 44.7% | 51.2% | 57.1%
    2018 I 47.2% | 8.7% | 22.7% | 76.6% | 33.2% | 41.1% | 45.9% | 52.9% | 59.0%
    2019 I 48.0% I 9.3% I 19.1% I 74.1% I 36.2% I 43.2% I 47.9% I 53.0% I 60.0%

  - Medicaid

    YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH
    2017 I 54.5% | 12.9% | 0.0% | 76.9% | 37.3% | 48.2% | 55.9% | 63.7% | 69.5%
    2018 I 55.0% | 11.8% | 20.4% | 76.7% | 33.1% | 41.7% | 50.0% | 63.7% | 69.1%
    2019 I 54.6% I 12.0% I 23.8% I 100.0% I 39.0% I 46.4% I 55.5% I 62.7% I 71.2%

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

- None identified

Potential harms

- None identified

Additional Feedback: N/A

***Questions for the Committee**:*

- Do the performance results represent improvement? Is this a concern?
- Do the benefits of the measure outweigh any potential unintended consequences?

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

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

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

- Benefits outweigh harms and presumably better follow up will result in improved outcomes (if you believe the widespread diagnosis of ADHD and medication and psychotherapeutics) is appropriate.

- This is only measuring follow up after initiation of prescribing. It does not measure effectiveness of treatment or outcomes that matter to patients and families.
- No concerns.
- Benefits outway harms.
- No problems in this area.
- Overall usable. One possible unintended consequence is that children who are not able to get into the office in person would not be prescribed this medication perpetuating the disparity gap.
- I believe this would be useful
- Agree regular followup is necessary. but i still would like to see more evidence to support the 1 month requirement. the developer
- States there is no new information to contradict it; but i would like to see more evidence to support the 30 day requirement. i would appreciate it if the developer can elaborate more about where in the clinical guidelines does it specify 30 days.
- By allowing to count a telephone contact during C& M phase, questions of adequate medication safety monitoring are raised.
- No concerns.
- Performance results could definitely improve care.

Criterion 5: Related and Competing Measures

Related or competing measures

- Developer did not identify any related or competing measures
- NQF staff identified the following measures as related, but not competing
    - 0106: Diagnosis of ADHD in primary care for school age children and adolescents
    - 0107: Management of ADHD in primary care for school age children and adolescents

Harmonization

NQF staff determined that these measures are harmonized to the extent possible.

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

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

- I want to know more about the current measures.
- I am not aware of any.
- Related measures are harmonized.
- Related but not competing.
- Nothing significant.
- None.
- No.
- No competing.
- There are related but not competing measures.

Public and Member Comments

No Comments and Member Support/Non-Support Submitted as of: 06/05/2020

Staff Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0108

Measure Title: Follow-Up Care for Children Prescribed ADHD Medication (ADD)

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: Specifications, 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 identified

RELIABILITY: TESTING

Submission document: Specifications, testing attachment questions 1.1-1.4 and section 2a2

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

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   - Developer performed reliability testing using a signal-to-noise analysis with the beta-binomial methodology.
   - This is a common approach used for reliability testing at the score level for measures of pass/fail events.

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

   Submission document: Testing attachment, section 2a2.3

   - Developer performed reliability testing using a signal-to-noise analysis with the beta-binomial methodology.
   - This is a common approach used for reliability testing at the score level for measures of pass/fail events.
   - While the developer has only reported a single summary statistic, developer has met with NQF staff to request the opportunity to share a new approach to include confidence intervals around score level reliability point estimates prior to the Committee meeting.
     - Developer was referred to NQF submission criteria: "For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred." (NQF Measure Evaluation Criteria, pg. 18)
   - Developer offered the table below as the results:

Measure Rate	Signal-to-Noise Reliability	
	Commercial	Medicaid
Initiation Phase	0.88	0.98
Continuation and Maintenance Phase	0.74	0.94

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.**

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    - No concerns

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

    Submission document: Testing attachment, section 2b4.

    - Performing a t-test of randomly-selected plans at the 25th and 75th performance percentile is an appropriate approach.

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.

    - No concerns

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

    Submission document: Testing attachment, section 2b6.

    - Developer describes audit process for ensuring that plans do not have missing data; no concerns.

16. **Risk Adjustment**

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

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

    ☐ Yes ☐ No ☒ Not applicable

    16c. **Social risk adjustment:**

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

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

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

    16d.**Risk adjustment summary:**

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

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

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

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

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

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

    - No risk adjustment for this process measure

VALIDITY: TESTING

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

    ☒ Face validity

    ☒ Empirical validity testing of the measure score

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

    Submission document: Testing attachment, section 2b2.2

    - Developer offered two forms of testing
        - Face validity
        - Construct validity using correlation statistics between the individual rates of the measures, as well as with a measure in a comparable quality domain
    - Both approaches are commonly used and appropriate forms of score-level testing
20. **Assess the results(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.3

    - Face validity results found the measure to be sound
    - Statistical results from Pearson's Correlation analysis comparing the measure's two rates:
        - Commercial: 0.78
        - Medicaid: 0.89
    - Statistical results from Pearson's Correlation analysis comparing the measure with *Use of First-line Psychosocial Care for Children and Adolescents on Antipsychotics*:
        - Commercial: Initiation: 0.26; Continuation and Maintenance: 0.14
        - Medicaid: Initiation: 0.31; Continuation and Maintenance: 0.30
    - The individual rates of the measures were highly correlated. Correlation between the measure rates and the comparator measures were weak to moderate.
21. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (score-level testing was not performed)
22. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**

    NOTE that data element validation from the literature is acceptable.

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

    ☐ **Yes**

    ☐ **No**

    ☒ **Not applicable** (data element testing was not performed)
23. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

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

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

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

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

24. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**
    - Results were fairly good, though there was not an especially strong correlation between the measure and an outside measure.
    - No concerns.

ADDITIONAL RECOMMENDATIONS

If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing

- Developer should be asked to clarify their reliability results, especially for the commercial plans. The spread of the data is not clear, nor what is meant by "overall reliability".

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

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

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

nqf_evidence_attachment_7.1.docx

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

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

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0108

Measure Title: Follow-Up Care for Children Prescribed ADHD Medication (ADD)

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

Date of Submission: 4/2/2020

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

Outcome

☐ Outcome: Click here to name the health outcome

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

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

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

☒ Process: Follow-Up Care for Children Prescribed ADHD Medication

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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

Children newly prescribed attention-deficit hyperactivity disorder (ADHD) medication >> timely follow-up visits occur >> medication effectiveness and any adverse effects are assessed >> dose is adjusted if needed >> treatment adherence and health outcomes are improved

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.)

NA.

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

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

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Table 1: American Academy of Pediatrics Guidelines	
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	• ADHD: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents • American Academy of Pediatrics (AAP) • 2019 • ADHD: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. Subcommittee On Attention-Deficit/Hyperactivity Disorder. September 30, 2019, peds.2019;144; DOI: 10.1542/peds.2019-2528 • https://pediatrics.aappublications.org/content/pediatrics/144/4/e20192528.full.pdf

Table 1: American Academy of Pediatrics Guidelines	
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.	American Academy of Pediatrics Clinical Practice Guideline for the Diagnosis, Evaluation and Treatment of ADHD in Children and Adolescents *Key Action Statement (KAS) 1:* The pediatrician or other primary care clinicians (PCC) should initiate an evaluation for ADHD for any child or adolescent age 4 years to the 18th birthday who presents with academic or behavioral problems and symptoms of inattention, hyperactivity, or impulsivity. *Grade B: Strong Recommendation* *KAS 4:* ADHD is a chronic condition; therefore, the PCC should manage children and adolescents with ADHD in the same manner that they would children and youth with special health care needs, following the principles of the chronic care model and the medical home. *Grade B: Strong Recommendation* *KAS 5b*: For elementary and middle school-aged children (age 6 years to the 12th birthday) with ADHD, the PCC should prescribe FDA-approved medications for ADHD, along with parent training in behavior management (PTBM) and/or behavioral classroom intervention (preferably both PTBM and behavioral classroom interventions). Educational interventions and individualized instructional supports, including school environment, class placement, instructional placement, and behavioral supports, are a necessary part of any treatment plan and often include an IEP or a rehabilitation plan (504 plan). *Grade A: Strong Recommendation*
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The grade assigned by the AAP to the evidence supporting the listed Clinical Practice Guidelines for the Diagnosis, Evaluation and Treatment of ADHD in Children and Adolescents were A and B. Grade A: Meta-analysis, systematic reviews of RCTs or diagnostic studies with minor limitations; overwhelmingly consistent evidence from observational studies Grade B: RCTs or diagnostic studies with minor limitations; overwhelmingly consistent evidence from observational studies

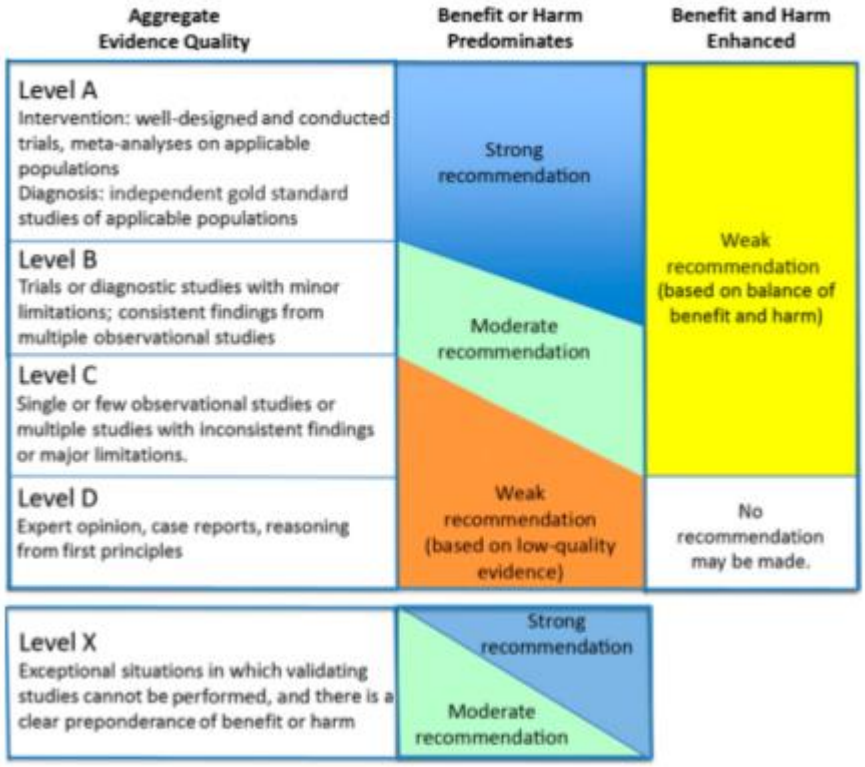
Table 1: American Academy of Pediatrics Guidelines	
Provide all other grades and definitions from the evidence grading system	 FIGURE 1 AAP rating of evidence and recommendations. Figure above explains the quality of evidence indicated by the grading system used, and how strong of a recommendation is warranted based on the evidence provided at each grading level.
Grade assigned to the **recommendation** with definition of the grade	The AAP assigned the categorization of *Strong Recommendation* to the listed Clinical Practice Guidelines for the Diagnosis, Evaluation and Treatment of ADHD in Children and Adolescents. Strong Recommendation: A strong recommendation means that the committee believes that the benefits of the recommended approach clearly exceed the harms of that approach (or, in the case of a strong negative recommendation, that the harms clearly exceed the benefits) and that the quality of the evidence supporting this approach is either excellent or impossible to obtain. Clinicians should follow such guidance unless a clear and compelling rationale for acting in a contrary manner is present
Provide all other grades and definitions from the recommendation grading system	**American Academy of Pediatrics Grading System** • *Recommendation*: A recommendation means that the committee believes that the benefits exceed the harms (or, in the case of a negative recommendation, that the harms exceed the benefits), but the quality of the evidence on which this recommendation is based is not as strong. Clinicians also generally should follow such guidance but also should be alert to new information and sensitive to patient preferences • *Option*: An option means either that the evidence quality that exists is suspect or that well-designed, well-conducted studies have demonstrated little clear advantage to one approach versus another. Options offer clinicians flexibility in their decision-making regarding appropriate practice, although they may set boundaries on

Table 1: American Academy of Pediatrics Guidelines	
	alternatives. Patient preference should have a substantial role in influencing clinical decision-making, particularly when policies are expressed as options. No Recommendation: No recommendation is made when there is both a lack of pertinent evidence and an unclear balance between benefits and harms. Clinicians should feel little constraint in their decision-making when addressing areas with insufficient evidence. Patient preference should have a substantial role in influencing clinical decision-making.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Guidelines from the American Academy of Pediatrics (Wolraich et al. 2011) cite a randomized control trial of 600 children diagnosed with ADHD ages 7-9 years old, as well as a prospective observational cohort study of 34 children. It also cites two systematic literature reviews and the chronic care model (Bodenheimer, Wagner, & Grumbach 2002). The action statement also received a grade of B, which indicates high quality evidence including randomized controlled trials. The evidence supporting the AAP guidelines received a grade of B, indicating that the guideline is supported by strong evidence consisting of randomized controlled trials and that there is a preponderance of benefit compared to harm. Additionally, the AAP guideline is based on 15 randomized controlled trials in addition to multiple controlled and uncontrolled trials, all providing evidence of the efficacy of continuous medication treatment and exploring side effects and other aspects of ADHD medication use that require monitoring.
Estimates of benefit and consistency across studies	The evidence supporting the guidelines demonstrate the benefits of consistent treatment, side-effect monitoring and medication adjustment for children and adolescents with ADHD. Timely follow-up visits ensure children and adolescents on AHDH medications receive these services. Both the AAP and AACAP guidelines are based on the Multi-Modal Treatment Study of Children with ADHD (MTA) (in addition to other studies). In the MTA study, children with ADHD were randomized to four groups: algorithmic medication treatment alone, psychosocial treatment alone, a combination of algorithmic medication management and psychosocial treatment, and community treatment. Algorithmic medication treatment consisted of monthly appointments in which the dose of medication was titrated according to parent and teacher rating scales. Children in all four treatment groups demonstrated benefits to treatment in terms of reduced symptoms of ADHD compared to baseline. The two groups that received algorithmic medication management showed a superior outcome with regard to ADHD symptoms compared with those that received intensive behavioral treatment alone or community treatment (MTA Cooperative Group, 1999a [rct]). Once the study treatments ceased at 14 months, the combined and medication groups lost some of their treatment gains, in part because of medication discontinuation and in part because the medication was now being given in the community with less careful monitoring and dose adjustment (MTA Cooperative Group, 2004a [rct], 2004b [rct]).
What harms were identified?	The American Academy of Pediatrics provided an analysis of net benefit and concluded that there is a preponderance of benefit over harm because of the opportunity to assess adverse effects of medication and to sustain treatment.
Identify any new studies conducted since the SR. Do the new studies	There have been no new studies that contradict the current body of evidence.

Table 1: American Academy of Pediatrics Guidelines	
change the conclusions from the SR?	

Table 2: American Academy of Child and Adolescent Psychiatry Guidelines	
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	Practice parameter for the assessment and treatment of children and adolescents with attention-deficit/hyperactivity disorder. American Academy of Child and Adolescent Psychiatry (AACAP) 2007 Practice Parameter for the Assessment and Treatment of Children and Adolescents With Attention-Deficit/Hyperactivity Disorder Pliszka, Steven. Journal of the American Academy of Child & Adolescent Psychiatry , Volume 46 , Issue 7 , 894 – 921. http://www.jaacap.com/article/S0890-8567(09)62182-1/abstract
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.	Overall Guideline The key to effective long-term management of the patient with ADHD is continuity of care with a clinician experienced in the treatment of ADHD. The frequency and duration of follow-up sessions should be individualized for each family and patient, depending on the severity of ADHD symptoms; the degree of comorbidity of other psychiatric illness; the response to treatment; and the degree of impairment in home, school, work, or peer-related activities. The clinician should establish an effective mechanism for receiving feedback from the family and other important informants in the patient's environment to be sure symptoms are well controlled and side effects are minimal. Although this parameter does not seek to set a formula for the method of follow-up, significant contact with the clinician should typically occur two to four times per year in cases of uncomplicated ADHD and up to weekly sessions at times of severe dysfunction or complications of treatment. Specific Recommendations Recommendation 6: A Well-Thought-Out and Comprehensive Treatment Plan Should Be Developed for the Patient With ADHD. The treatment plan should be reviewed regularly and modified if the patient's symptoms do not respond. Minimal Standard [MS] Recommendation 9. During a Psychopharmacological Intervention for ADHD, the Patient Should Be Monitored for Treatment-Emergent Side Effects. Minimal Standard [MS] Recommendation 12. Patients Should Be Assessed Periodically to Determine Whether There Is Continued Need for Treatment or If Symptoms Have Remitted. Treatment of ADHD Should Continue as Long as Symptoms Remain Present and Cause Impairment. Minimal Standard [MS]

Table 2: American Academy of Child and Adolescent Psychiatry Guidelines	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The grade assigned by AACAP to the evidence supporting the listed Practice Parameters for the Assessment and Treatment of Children and Adolescents with ADHD varied by study and included [rct] and [ut]: • [rct] Randomized, controlled trial is applied to studies in which subjects are randomly assigned to two or more treatment conditions. • [ut] Uncontrolled trial is applied to studies in which subjects are assigned to one treatment condition
Provide all other grades and definitions from the evidence grading system	The grades assigned by the AACAP to evidence supporting the Practice Parameters for the Assessment and Treatment of Children and Adolescents with ADHD noted the strength of the study by listing the study type. Other studies used to support the guidelines included: • [rct] Randomized, controlled trial is applied to studies in which subjects are randomly assigned to two or more treatment conditions. • [ct] Controlled trial is applied to studies in which subjects are nonrandomly assigned to two or more treatment conditions. • [ut] Uncontrolled trial is applied to studies in which subjects are assigned to one treatment condition. • [cs] Case series/report is applied to a case series or a case report.
Grade assigned to the **recommendation** with definition of the grade	AACAP assigned a grade of [MS] Minimal Standard to the listed Practice Parameters for the Assessment and Treatment of Children and Adolescents with ADHD. [MS] Minimal Standard is applied to recommendations that are based on rigorous empirical evidence (e.g., randomized, controlled trials) and/or overwhelming clinical consensus. Minimal standards apply more than 95% of the time (i.e., in almost all cases).
Provide all other grades and definitions from the recommendation grading system	American Academy of Child and Adolescent Psychiatry Grading System [CG] Clinical Guideline is applied to recommendations that are based on strong empirical evidence (e.g., nonrandomized, controlled trials) and/or strong clinical consensus. Clinical guidelines apply approximately 75% of the time (i.e., in most cases). [OP] Option is applied to recommendations that are acceptable based on emerging empirical evidence (e.g., uncontrolled trials or case series/reports) or clinical opinion, but lack strong empirical evidence and/or strong clinical consensus. [NE] Not Endorsed is applied to practices that are known to be ineffective or contraindicated.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Guidelines from the American Academy of Child and Adolescent Psychiatry (Pliszka 2007) cite two randomized control trials, one of which enrolled 600 children diagnosed with ADHD ages 7-9 years. The other study enrolled 103 children diagnosed with ADHD also ages 7-9. In addition to these large randomized control trials, each recommendation cited additional studies which provided further evidence examining treatment planning for children with ADHD, side effects associated with ADHD medications, medication adherence, and treatment adjustment.

Table 2: American Academy of Child and Adolescent Psychiatry Guidelines	
	Recommendation 6: Treatment Plan review and modification: 3 RCTs Recommendation 9: Side Effect Monitoring: 4 RCTS, 1 CT, 1 UT Recommendation 12: Periodic Assessment of Symptoms: 8 RCTs, 5 UTs Overall, the quality of the evidence regarding follow-up care for children with a prescription for an ADHD medication is good. Guidelines from the American Academy of Pediatrics and the American Academy of Child and Adolescent Psychiatry cite the Multi-Modal Treatment Study of Children with ADHD (MTA). The 1999-published MTA study, sponsored by the National Institute of Mental Health, was a randomized control trial, multi-site study of nearly 600 elementary school children, 7-9 years of age who were diagnosed with ADHD and randomly assigned to one of four treatment modes: medication alone; psychosocial/behavioral treatment alone; a combination of both; or routine community care. The MTA study demonstrated that, on average, carefully monitored medication management with monthly follow-up is more effective than intensive behavioral treatment for ADHD symptoms, for periods lasting as long as 14 months. The quality of this study can be considered high due to the randomization and large sample size. The AACAP guideline is based on 15 randomized controlled trials in addition to multiple controlled and uncontrolled trials, all providing evidence of the efficacy of continuous medication treatment and exploring side effects and other aspects of ADHD medication use that require monitoring.
Estimates of benefit and consistency across studies	The evidence supporting the guidelines demonstrate the benefits of consistent treatment, side-effect monitoring and medication adjustment for children and adolescents with ADHD. Timely follow-up visits ensure children and adolescents on AHDH medications receive these services. Both the AAP and AACAP guidelines are based on the Multi-Modal Treatment Study of Children with ADHD (MTA) (in addition to other studies). In the MTA study, children with ADHD were randomized to four groups: algorithmic medication treatment alone, psychosocial treatment alone, a combination of algorithmic medication management and psychosocial treatment, and community treatment. Algorithmic medication treatment consisted of monthly appointments in which the dose of medication was titrated according to parent and teacher rating scales. Children in all four treatment groups demonstrated benefits to treatment in terms of reduced symptoms of ADHD compared to baseline. The two groups that received algorithmic medication management showed a superior outcome with regard to ADHD symptoms compared with those that received intensive behavioral treatment alone or community treatment (MTA Cooperative Group, 1999a [rct]). Once the study treatments ceased at 14 months, the combined and medication groups lost some of their treatment gains, in part because of medication discontinuation and in part because the medication was now being given in the community with less careful monitoring and dose adjustment (MTA Cooperative Group, 2004a [rct], 2004b [rct]). In terms of side-effect monitoring, the AACAP guidelines are based on four randomized controlled trials, one controlled trial and one uncontrolled trial. These trials found that it is prudent to monitor side effects in order to optimize patient outcomes.

Table 2: American Academy of Child and Adolescent Psychiatry Guidelines	
What harms were identified?	N/A
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Numerous (>100) studies related to the care for patients with ADHD have been published since the publication of this guideline, none of which contradict the need for appropriate follow-up once treatment with medication begins.

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

1b. Performance Gap

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

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

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

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

Attention-deficit/hyperactivity disorder (ADHD) is a brain disorder marked by an ongoing pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development. Medications can improve function, but proper monitoring is recommended. The intent of this measure is to ensure timely and continuous follow-up visits for children who are newly prescribed ADHD medication. The goal is to encourage monitoring of children for medication effectiveness, occurrence of side effects and adherence.

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

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

The following data demonstrate room for improvement among health plans.

These rates are extracted from HEDIS data collection and reflect the most recent years of measurement for this measure. For HEDIS 2019 (calendar year 2018), HEDIS measures covered 116 million commercial health plan members and 54 million Medicaid enrollees. Data are summarized at the health plan level and stratified by product line (i.e. commercial, Medicaid).

2017 -2019 Performance Data

INITIATION PHASE

Commercial

YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH

2017 I 39.5% | 8.2% | 11.3% | 75.6% | 30.5% | 35.0% | 38.6% | 43.6% | 50.0%

2018 I 40.7% | 8.1% | 22.2% | 75.6% | 31.2% | 35.9% | 41.2% | 44.3% | 51.8%

2019 I 40.0% I 8.3% I 14.0% I 70.5% I 29.8% I 35.0% I 40.0% I 44.3% I 50.4%

Medicaid

YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH

2017 I 44.5% | 10.4% | 0.0% | 86.1% | 31.7% | 39.0% | 44.8% | 51.8% | 57.1%

2018 I 44.6% | 9.1% | 20.4% | 73.4% | 34.3% | 38.2% | 45.0% | 50.8% | 55.9%

2019 I 44.2% I 9.7% I 22.5% I 92.5% I 33.9% I 37.9% I 43.4% I 49.9% I 56.6%

CONTINUATION AND MAINTENANCE PHASE

Commercial

YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH

2017 I 46.1% | 8.6% | 25.7% | 74.0% | 36.4% | 40.3% | 44.7% | 51.2% | 57.1%

2018 I 47.2% | 8.7% | 22.7% | 76.6% | 33.2% | 41.1% | 45.9% | 52.9% | 59.0%

2019 I 48.0% I 9.3% I 19.1% I 74.1% I 36.2% I 43.2% I 47.9% I 53.0% I 60.0%

Medicaid

YEAR| MEAN | ST DEV | MIN | MAX | 10TH | 25TH | 50TH | 75TH | 90TH

2017 I 54.5% | 12.9% | 0.0% | 76.9% | 37.3% | 48.2% | 55.9% | 63.7% | 69.5%

2018 I 55.0% | 11.8% | 20.4% | 76.7% | 33.1% | 41.7% | 50.0% | 63.7% | 69.1%

2019 I 54.6% I 12.0% I 23.8% I 100.0% I 39.0% I 46.4% I 55.5% I 62.7% I 71.2%

2017 -2019 Denominator Data

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.

INITIATION PHASE

Commercial

YEAR| N PLANS | Mean Denominator Size per plan

2017 I 343 | 403

2018 I 325 | 415

2019 I 319 I 417

Medicaid

YEAR| N PLANS | Mean Denominator Size per plan

2017 I 196 | 1,244

2018 I 192 | 1,339

2019 I 183 I 1,466

CONTINUATION AND MAINTENANCE PHASE

Commercial

YEAR| N PLANS | Mean Denominator Size per plan

2017 I 214 | 163

2018 I 203 |165

2019 I 205 I 162

Medicaid

YEAR| N PLANS | Mean Denominator Size per plan

2017 I 175 | 305

2018 I 173 | 334

2019 I 169 I 340

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

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

Studies suggest children from minority families experience decreased access to and utilization of health services for ADHD, even after controlling for poverty and health insurance status (Miller, Nigg, & Miller 2009; Morley, 2011; Nasol et al., 2018). Although the prevalence of ADHD in minority children is believed to be equal to or even greater than the prevalence in non-minority children, studies indicate that rates of both diagnosis and treatment of ADHD are much lower among minority children compared to non-minority children. Specifically, children who are black, are raised in primarily non-English speaking households, have limited access to the health care system, and are poorer (Bailey et al. 2014; Morgan et al. 2014; Flores & Lin 2013; Froehlich et al. 2007). A 2016 study by Coker et al. found robust disparities in ADHD diagnoses and patient-reported medication use for youth enrolled in fifth through tenth grade, after controlling for sociodemographic factors, ADHD symptoms, and mental health comorbidities. In comparison to white children, Latino and African American children were less likely to have an ADHD diagnosis and to be taking ADHD medication. Further studies also indicate that, among children with ADHD, Hispanic and African American children were less often reported to use medication than white children (Bailey et al. 2014; Pastor & Reuben 2005). The NIMH Multisite Multimodal Treatment Study of Children with Attention-Deficit/Hyperactivity Disorder cited by American Academy of Child and Adolescent Psychiatry (AACAP) and American Academy of Pediatrics (AAP) indicates that certain disparities affected eight-year prospective follow-up. Participants lost to follow-up were “more often male, had younger mothers, had less educated parents, had lower parent income, and were more likely to have been on welfare at baseline” (Brooke et al. 2009). African American and Latino youth also discontinue medication and treatment at disproportionate rates (Cummings et al., 2017). A 2017 study of Medicaid enrolled children ages 6-12 found that compared to their white counterparts, “African American and Hispanic children were 22.4% and 16.7% points more likely to discontinue medication, and 13.1% and 9.4% points more likely to disengage from treatment” (Cummings et al. 2017). Studies suggest effective ADHD treatment in minority children may be affected by cultural norms surrounding ADHD. For example, some minority communities perceive that mental illness is a sign of personal weakness or that seeking treatment will jeopardize future employment or military service (Bailey et al. 2014). These perceptions lead to a lack of treatment seeking by these individuals and lack of appropriate screening (Price et al., 2013). Cost can also be a barrier to adequate care, since across racial and ethnic groups, an estimated 44% of school-aged children with a diagnosis of ADHD report an adverse family financial impact, and 11% indicate unmet need for ADHD treatment (Nasol et al. 2018.

Bailey, R. K., Jaquez-Gutierrez, M. C., & Madhoo, M. 2014. “Sociocultural issues in african American and Hispanic minorities seeking care for attention-deficit/hyperactivity disorder”. The Primary Care Companion for CNS Disorders 16(4).

Brooke S.G. Molina Ph.D., Stephen P. Hinshaw Ph.D., James M. Swanson Ph.D., L. Eugene Arnold M.D., M.Ed., Benedetto Vitiello M.D., Peter S. Jensen M.D., Jeffery N. Epstein Ph.D., Betsy Hoza Ph.D., Lily Hechtman M.D., Howard B. Abikoff Ph.D., Glen R. Elliott Ph.D., M.D., Laurence L. Greenhill M.D., Jeffrey H. Newcorn M.D., Karen C. Wells Ph.D., Timothy Wigal Ph.D., Robert D. Gibbons Ph.D., Kwan Hur Ph.D. and Patricia R. Houck M.S. 2009. “The MTA at 8 Years: Prospective Follow-up of Children Treated for Combined-Type ADHD in a Multisite Study.” Journal of the American Academy of Child and Adolescent Psychiatry 48(5):484-500.

Coker T., Elliot M., Toomey S., Schwebel D., Cuccaro P., Emery S., Davies S. Visser S., Schuster M. 2016. " Racial and Ethnic Disparities in ADHD Diagnosis and Treatment" . Pediatrics 138(3).

Cummings, J. R., Ji, X., Allen, L., Lally, C., & Druss, B. G. 2017. Racial and Ethnic Differences in ADHD Treatment Quality Among Medicaid-Enrolled Youth. Pediatrics, 139(6), e20162444. https://doi.org/10.1542/peds.2016-2444

Flores G. & H. Lin. 2013. "Trends in racial/ethnic disparities in medical and oral health, access to care, and use of services in US children: has anything changed over the years?" International Journal for Equity in Health 12:10.

Froehlich T.E., B.P. Lanphear, J.N. Epstein. 2007. "Prevalence, Recognition, and Treatment of Attention-Deficit/Hyperactivity Disorder in a National Sample of US Children." Archives of Pediatric and Adolescent Medicine 161(9):857-64.

Miller T.W., J.T. Nigg, R.L. Miller. 2009. "Attention deficit hyperactivity disorder in African American children: what can be concluded from the past ten years?" Clinical Psychological Review 29(1):77-86.

Morley, C. P. 2010. Disparities in Adhd Assessment, Diagnosis, and Treatment. The International Journal of Psychiatry in Medicine, 40(4), 383–389. https://doi.org/10.2190/PM.40.4.b

Morgan P.L., M.M. Hillemeier, G. Farkas, S. Maczuga. 2014. "Racial/ethnic disparities in ADHD diagnosis by kindergarten entry." Journal of Child Psychology and Psychiatry, and Allied Disciplines 55(8):905-13.

Nasol E., Lindly O., Cheavez A., Zuckerman K. 2018. "Unmet Need and Financial Impact Disparities for US Children with ADHD.

Pastor P.N. & C.A. Reuben. 2005. "Racial and ethnic differences in ADHD and LD in young school-age children: parental reports in the National Health Interview Survey." Public Health Reports 120(4): 383–392.

Price, J. H., Khubchandani, J., McKinney, M., & Braun, R. 2013. "Racial/ethnic disparities in chronic diseases of youths and access to health care in the United States". BioMed Research International.

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):*

Behavioral Health, Behavioral Health : Attention Deficit Hyperactivity Disorder (ADHD)

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

Access to Care, Person-and Family-Centered Care

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.)*

NA

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:**

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.

Clarified that for the Continuation and Maintenance Phase Rate, visits must be on different dates of service.

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).

Among children newly prescribed ADHD medication, those who had timely and continuous follow-up visits.

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).

RATE 1. INITIATION PHASE NUMERATOR

An outpatient, intensive outpatient or partial hospitalization follow-up visit with a practitioner with prescribing authority, within 30 days after the earliest prescription dispensing date for a new ADHD medication. Any of the following code combinations billed by a practitioner with prescribing authority meet criteria:

- An outpatient visit (Visit Setting Unspecified Value Set with Outpatient POS Value Set).
- An outpatient visit (BH Outpatient Value Set).
- An observation visit (Observation Value Set).
- A health and behavior assessment/intervention (Health and Behavior Assessment/Intervention Value Set).
- An intensive outpatient encounter or partial hospitalization (Visit Setting Unspecified Value Set with Partial Hospitalization POS Value Set).
- An intensive outpatient encounter or partial hospitalization (Partial Hospitalization/Intensive Outpatient Value Set).
- A community mental health center visit (Visit Setting Unspecified Value Set with Community Mental Health Center POS Value Set).

Note: Do not count a visit on the Index Prescription Start Date as the Initiation Phase visit. Do not count visits billed with a telehealth modifier (Telehealth Modifier Value Set) or billed with a telehealth POS code (Telehealth POS Value Set).

RATE 2. CONTINUATION AND MAINTENANCE PHASE NUMERATOR

Children who are numerator compliant for Rate 1. Initiation Phase, AND have documentation of at least two follow-up visits on different dates of service with any practitioner from 31–300 days (9 months) after the earliest prescription dispensing date for a new ADHD medication.

One of the two visits (during days 31–300) may be a telephone visit (Telephone Visits Value Set) with any practitioner. Identify follow-up visits using the code combinations below, then identify telehealth visits by the presence of a telehealth modifier (Telehealth Modifier Value Set) or the presence of a telehealth POS code (Telehealth POS Value Set) on the claim.

Any of the following code combinations identify follow-up visits:

• An outpatient visit (Visit Setting Unspecified Value Set with Outpatient POS Value Set).

• An outpatient visit (BH Outpatient Value Set).

• An observation visit (Observation Visit Value Set).

• A health and behavior assessment/intervention (Health and Behavior Assessment/Intervention Value Set).

• An intensive outpatient encounter or partial hospitalization (Visit Setting Unspecified Value Set with Partial Hospitalization POS Value Set).

• An intensive outpatient encounter or partial hospitalization (Partial Hospitalization/Intensive Outpatient Value Set).

• A community mental health center visit (Visit Setting Unspecified Value Set with Community Mental Health Center POS Value Set).

• A telehealth visit (Visit Setting Unspecified Value Set with Telehealth POS Value Set).

• A telephone visit (Telephone Visits Value Set).

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

Children 6-12 years of age newly prescribed ADHD medication.

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).

RATE 1. INITIATION PHASE DENOMINATOR

Children age 6 as of March 1 of the measurement year; 12 years as of February 28 of the measurement year. who were dispensed a new ADHD medication during the 12-month Intake Period (Table ADD-A). Patients must have all of the following:(1) A 120-day (4-month) negative medication history on or before the Index Prescription Date. The Index Prescription Start Date is the dispensing date of the earliest ADHD prescription in the Intake Period with a Negative Medication History.

(2) Continuous enrollment for 120 days prior to the Index Prescription Start Date through 30 days after the Index Prescription Start Date.

(3) Exclude patients who had an acute inpatient encounter for mental health or chemical dependency during the 30 days after the Index Prescription Start Date. An acute inpatient encounter in combination with any of the following meet criteria:

A principal mental health diagnosis (Mental Health Diagnosis Value Set).

A principal diagnosis of chemical dependency (Chemical Dependency Value Set)

Due to the extensive volume of codes associated with identifying the denominator for this measure, we are attaching a separate file with code value sets. See code value sets located in question S.2b.

ADHD MEDICATIONS LIST

CNS stimulants: Amphetamine-dextroamphetamine, dexmethylphenidate, dextroamphetamine, lisdexamfetamine, methamphetamine, methylphenidate

Alpha-2 receptor agonists: Clonidine, guanfacine

Miscellaneous: Atomoxetine

RATE 2. CONTINUATION AND MAINTENANCE PHASE DENOMINATOR

Children who meet the eligible population criteria for Rate 1. Initiation Phase who have been continuously enrolled in the organization for 120 days (4 months) prior to the Index Prescription Start Date and 300 days (10 months) after the Index Prescription Start Date. Patients must have all of the following:

(1) The patient must have filled a sufficient number of prescriptions to provide continuous treatment for at least 210 days out of the 300-day period after the Index Prescription Start Date. The definition of "continuous medication treatment" allows gaps in medication treatment, up to a total of 90 days during the 300-day (10-month) period. (This period spans the Initiation Phase [1 month] and the C&M Phase [9 months].)

Gaps can include either washout period gaps to change medication or treatment gaps to refill the same medication.

Regardless of the number of gaps, the total gap days may be no more than 90. The organization should count any combination of gaps (e.g., one washout gap of 14 days and numerous weekend drug holidays).

(2) Exclude patients who had an acute inpatient encounter for mental health or chemical dependency during the 300 days (10 months) after the Index Prescription Start Date. An acute inpatient encounter in combination with any of the following meet criteria:

A principal mental health diagnosis (Mental Health Diagnosis Value Set).

A principal diagnosis of chemical dependency (Chemical Dependency Value Set).

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

Children who had an acute inpatient encounter for mental health or chemical dependency following the Index Prescription Start Date

Children with a diagnosis of narcolepsy: Many of the medications used to identify patients for the denominator of this measure are also used to treat narcolepsy. Children with narcolepsy who are pulled into the denominator are then removed by the narcolepsy exclusion.

Children using hospice services during the measurement year. Children in hospice may not be able to receive the necessary follow-up care.

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.)*

2016 Submission

Exclude from the denominator for both rates, children who had an acute inpatient encounter for mental health or chemical dependency during the 30 days after the Index Prescription Start Date

Exclude from the denominator for both rates, children with a diagnosis of narcolepsy (Narcolepsy Value Set) any time during their history through December 31 of the measurement year

Exclude from the denominator for both rates, patients who use hospice services or elect to use a hospice benefit any time during the measurement year, regardless of when the services began. These members may

be identified using various methods, which may include but are not limited to enrollment data, medical record or claims/encounter data

(Hospice Value Set).

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.)*

INITIATION PHASE: ELIGIBLE POPULATION

Step 1: Identify all children in the specified age range (Children 6-12 years of age: 6 as of March 1 of the measurement year; 12 years as of February 28 of the measurement year) who were dispensed an ADHD medication (ADHD MEDICATIONS LIST) during the 12-month Intake Period.

Step 2: Test for Negative Medication History. For each member identified in step 1, test each ADHD prescription for a Negative Medication History. The Index Prescription Start Date is the dispensing date of the earliest ADHD prescription in the Intake Period with a Negative Medication History.

Step 3: Calculate continuous enrollment. Patients must be continuously enrolled for 120 days (4 months) prior to the Index Prescription Start Date through 30 days after the Index Prescription Start Date.

Step 4: Exclude patients who had an acute inpatient encounter for mental health or chemical dependency during the 30 days after the Index Prescription Start Date. An acute inpatient encounter (Acute Inpatient Value Set) in combination with any of the following meet criteria: A principal mental health diagnosis (Mental Health Diagnosis Value Set) AND/OR A principal diagnosis of chemical dependency (Chemical Dependency Value Set).

Step 5: Determine the number of patients in the eligible population with an outpatient, intensive outpatient or partial hospitalization follow-up visit with a practitioner with prescribing authority, within 30 days after the Index Prescription Start Date. Any of the following code combinations billed by a practitioner with prescribing authority meet criteria:

- An outpatient visit (Visit Setting Unspecified Value Set with Outpatient POS Value Set).
- An outpatient visit (BH Outpatient Value Set).
- An observation visit (Observation Value Set).
- A health and behavior assessment/intervention (Health and Behavior Assessment/Intervention Value Set).

• An intensive outpatient encounter or partial hospitalization (Visit Setting Unspecified Value Set with Partial Hospitalization POS Value Set).

• An intensive outpatient encounter or partial hospitalization (Partial Hospitalization/Intensive Outpatient Value Set).

• A community mental health center visit (Visit Setting Unspecified Value Set with Community Mental Health Center POS Value Set).

Note: Do not count a visit on the Index Prescription Start Date as the Initiation Phase visit. Do not count visits billed with a telehealth modifier (Telehealth Modifier Value Set) or billed with a telehealth POS code (Telehealth POS Value Set).

Step 6: Calculate a rate (number of children receiving a follow-up visit with a prescriber within 30 days of the Index Prescription Start Date).

CONTINUATION AND MAINTENANCE PHASE: ELIGIBLE POPULATION

Step 1: Identify all patients who meet the eligible population criteria for Rate 1—Initiation Phase.

Step 2: Calculate continuous enrollment. Patients must be continuously enrolled in the organization for 120 days (4 months) prior to the Index Prescription Start Date and 300 days (10 months) after the Index Prescription Start Date.

Step 3: Calculate the continuous medication treatment. Using the patients in step 2, determine if the member filled a sufficient number of prescriptions to provide continuous treatment for at least 210 days out of the 300-day period after the Index Prescription Start Date. The definition of “continuous medication treatment” allows gaps in medication treatment, up to a total of 90 days during the 300-day (10-month) period. (This period spans the Initiation Phase [1 month] and the C&M Phase [9 months].) Gaps can include either washout period gaps to change medication or treatment gaps to refill the same medication. Regardless of the number of gaps, the total gap days may be no more than 90. The organization should count any combination of gaps (e.g., one washout gap of 14 days and numerous weekend drug holidays).

Step 4: Exclude patients who had an acute inpatient encounter for mental health or chemical dependency during the 300 days (10 months) after the Index Prescription Start Date. An acute inpatient encounter in combination with any of the following meet criteria:

A principal mental health diagnosis (Mental Health Diagnosis Value Set).

A principal diagnosis of chemical dependency (Chemical Dependency Value Set).

Step 5: Identify all patients in the eligible population who meet the following criteria:

(1) Numerator compliant for Rate 1—Initiation Phase, and

(2) At least two follow-up visits on different dates of service from 31–300 days (9 months) after the Index Prescription Start Date with any practitioner.

One of the two visits (during days 31–300) may be a telephone visit (Telephone Visits Value Set) with any practitioner. Any of the following code combinations identify follow-up visits:

• An outpatient visit (Visit Setting Unspecified Value Set with Outpatient POS Value Set).

• An outpatient visit (BH Outpatient Value Set).

• An observation visit (Observation Visit Value Set).

• A health and behavior assessment/intervention (Health and Behavior Assessment/Intervention Value Set).

• An intensive outpatient encounter or partial hospitalization (Visit Setting Unspecified Value Set with Partial Hospitalization POS Value Set).

• An intensive outpatient encounter or partial hospitalization (Partial Hospitalization/Intensive Outpatient Value Set).

• A community mental health center visit (Visit Setting Unspecified Value Set with Community Mental Health Center POS Value Set).

• A telehealth visit (Visit Setting Unspecified Value Set with Telehealth POS Value Set).

• A telephone visit (Telephone Visits Value Set).

Step 6: Calculate a rate (number of children receiving two follow-up visits with any practitioner from 31-300 days after the Index Prescription Start Date).

ADDITIONAL EXCLUSION:

Exclude from the denominator for both rates, patients with a diagnosis of narcolepsy (Narcolepsy Value Set) any time during their history through December 31 of the measurement year

NOTE

(1) Patients who have multiple overlapping prescriptions should count the overlap days once toward the days supply (whether the overlap is for the same drug or for a different drug).

(2) Organizations may have different methods for billing intensive outpatient encounters and partial hospitalizations. Some methods may be comparable to outpatient billing, with separate claims for each date of service; others may be comparable to inpatient billing, with an admission date, a discharge date and units of service. Organizations whose billing methods are comparable to inpatient billing may count each unit of service as an individual visit. The unit of service must have occurred during the period required for the rate (e.g., within 30 days after or from 31–300 days after the Index Prescription Start Date).

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

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

N/A

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

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

N/A

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

If other, please describe in S.18.

Claims

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

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

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

nqf_testing_attachment_7.1-637214197326050835.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*)**:** 0108
Measure Title: Follow-Up Care for Children Prescribed ADHD Medications
Date of Submission: 1/6/2020

Type of Measure:

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

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

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

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

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

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

2020 Submission

N/A

2016 Submission

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

2020 Submission

Testing of measure score reliability and validity was performed using data from calendar years 2016, 2017, and 2018.

2016 Submission

2014-2016

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

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

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

2020 Submission

This measure assesses whether children (age 6-12) newly prescribed attention-deficit/hyperactivity disorder (ADHD) medication receive at least three follow-up visits within a 10-month period, one of which occurring within 30 days of the first ADHD medication dispensed. This measure includes patients enrolled in commercial and Medicaid health plans. The first rate is for the Initiation Phase, assessing the percentage of members who had one follow-up visit with a prescribing practitioner within the 30 days of the first ADHD medication dispensed. The second rate is for the Continuation and Maintenance (C&M) Phase, assessing the percentage of members who remained on the medication for at least 210 days and had at least two follow-up visits with a practitioner within 9 months after the Initiation phase. The intended use of the measure is to assess the quality of care in health plans across the general child population. As required by the specified level of accountability, we conducted a field test with health plans to assess scientific acceptability, usability and feasibility and have subsequently gathered audited data from a large number of health plans.

Sample for measure score reliability testing and construct validity testing: The measure score reliability was calculated from HEDIS data that included 319 commercial health plans and 183 Medicaid health plans for the Initiation Phase and 205 commercial health plans and 169 Medicaid health plans for the Continuation and Maintenance Phase. The sample included all commercial and Medicaid health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size.

2016 Submission

Update: MEASURE SCORE RELIABILITY TESTING

The measure score reliability was calculated from HEDIS data that included 344 Commercial health plans and 180 Medicaid health plans for the Initiation Phase and 210 Commercial health plans and 151 Medicaid health plans for the Continuation and Maintenance Phase. The sample included all Commercial and Medicaid health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size.

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

2020 Submission

For HEDIS 2019 (calendar year 2018), HEDIS measures covered 116 million commercial health plan members and 54 million Medicaid enrollees. Data are summarized at the health plan level and stratified by product line (i.e. commercial, Medicare, Medicaid). Below is a description of the sample. It includes number of health plans included in HEDIS data collection and the average eligible population for the measure across health plans.

Table 1. Mean eligible population for the *Follow-Up Care for Children Prescribed ADHD Medication* measure by plan type, Initiation Phase, calendar year 2018 data

Product Type	Number of Plans	Mean number of eligible members per plan
Commercial	319	417
Medicaid	183	1,466

Table 2. Mean eligible population for the *Follow-Up Care for Children Prescribed ADHD Medication* measure by plan type, Continuation and Maintenance Phase, calendar year 2018 data

Product Type	Number of Plans	Mean number of eligible members per plan
Commercial	319	162
Medicaid	183	340

2016 Submission

2016 Update: MEASURE SCORE RELIABILITY TESTING

Patient data set for measure score reliability testing: In 2016, HEDIS measures covered 114.2 million commercial health plan beneficiaries, 47.0 million Medicaid beneficiaries, and 17.6 million Medicare beneficiaries. This measure applies to commercial and Medicaid plans. Data are summarized at the health plan level and stratified by product line. Below is a description of the sample, including number of health plans included and the median eligible population for the measure across health plans.

INITIATION PHASE

Product Line	Number of Plans	Mean number of eligible patients per plan
Commercial	344	397
Medicaid	180	1,160

CONTINUATION AND MAINTENANCE PHASE

Product Line	Number of Plans	Mean number of eligible patients per plan
Commercial	210	163
Medicaid	151	320

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

2020 Submission

No differences in the data used for reliability and construct validity testing.

2016 Submission

2016 Update: MEASURE SCORE RELIABILITY TESTING

Reliability of the measure score was tested using a beta-binomial calculation. This analysis included the entire HEDIS data sample (described 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.

2020 Submission

Social risk factor data were not available in reported results. This measure is specified to be reported separately by Medicaid and commercial plan types, which serves as a proxy for income and other socioeconomic factors.

2016 Submission

2016 Update: Measure performance was assessed by commercial, Medicaid and Medicare plan types.

2a2. RELIABILITY TESTING

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

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

2020 Submission

Reliability testing of performance measure score

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

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

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

2016 Submission

2016 Update: METHOD FOR MEASURE SCORE RELIABILITY TESTING

We used the beta binomial method as described below in our 2013 submission.

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

2020 Submission

Table 3 shows the estimated signal-to-noise reliability for each indicator.

Table 3. Estimated Signal-to-Noise Reliability for *Follow-Up Care for Children Prescribed ADHD Medication,* for Commercial and Medicaid Plans, calendar year 2018 data

Measure Rate	Signal-to-Noise Reliability	
	Commercial	Medicaid
Initiation Phase	0.88	0.98
Continuation and Maintenance Phase	0.74	0.94

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

2016 Submission

2016 Update: MEASURE SCORE RELIABILITY

Beta-Binomial Statistic For Each Measure Rate: Mean Reliability

Rate	*Commercial*	*Medicaid*
Initiation Phase	0.90	0.98
Continuation and Maintenance Phase	0.75	0.95

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

2020 Submission

The signal-to-noise reliability estimates are greater than 0.7, indicating the measure has very good reliability and provides confidence that one can distinguish the performance of one plan from another.

2016 Submission

2016 Update: INTERPRETATION OF RESULTS FOR MEASURE SCORE RELIABILITY TESTING

Beta binomial testing for this measure suggests the two rates (Initiation and Continuation and Maintenance) within this measure have strong reliability for commercial (0.90, 0.75) and Medicaid (0.98, 0.95) health plans.

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

2020 Submission

We assessed construct validity and face validity for this measure.

Method of testing construct validity

We tested for construct validity by exploring the following:

- Are the individual rates within the *Follow-Up Care for Children Prescribed ADHD Medication* measure correlated with one another.
- Is *Follow-Up Care for Children Prescribed ADHD Medication* correlated with the following HEDIS measure, *Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics*, which assesses the proportion of children and adolescents without a primary indication who had a new prescription for an antipsychotic medication and had documentation of psychosocial care as first-line treatment.

We hypothesized that rates within the *Follow-Up Care for Children Prescribed ADHD Medication* measure would be highly correlated, and that organizations that perform well on *Follow-Up Care for Children Prescribed ADHD Medication* should perform well on the other measure, *Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics*, given that they address the same or similar child populations and similar behavioral health conditions. To test these correlations, we used a Pearson correlation test. This test estimates the strength of the linear association between two variables. The magnitude of correlation ranges from -1 to +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

Method of assessing face validity

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

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

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

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA about proposed new measures. Public comment offers an opportunity to assess the validity, feasibility, importance and other attributes of a measure from a wider audience. For this measure, a majority of public

comment respondents supported the measure. NCQA MAPs and the technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. Face validity is then again systematically assessed by the CPM. The CPM reviews all comments before making a final decision and votes to recommend approval of new measures for HEDIS. NCQA's Board of Directors then approves new measures.

2016 Submission

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
2020 Submission

Statistical results of construct validity testing

Table 4a. Health-Plan Level Pearson Correlation Coefficients Among *Follow-Up Care for Children Prescribed ADHD Medication* Performance Scores Within Measure – **Commercial** Plans, calendar year 2018 data

Rate	Correlation Coefficient
	Initiation
Continuation and Maintenance	0.78*

*Significant at p<0.05

Table 4b. Health-Plan Level Pearson Correlation Coefficients Among *Follow-Up Care for Children Prescribed ADHD Medication* Performance Scores Within Measure – **Medicaid** Plans, calendar year 2018 data

Rate	Correlation Coefficient
	Initiation
Continuation and Maintenance	0.89*

*Significant at p<0.05

Table 5a. Health-Plan Level Pearson Correlation Coefficients Among *Follow-Up Care for Children Prescribed ADHD Medication* and *Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics* Measure Performance Scores – **Commercial** Plans, calendar year 2018 data

Rate	Correlation Coefficient
	Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics
Follow-Up Care for Children Prescribed ADHD Medication	
Initiation	0.26*
Continuation and Maintenance	0.14

*Significant at p<0.05

Table 5b. Health-Plan Level Pearson Correlation Coefficients Among *Follow-Up Care for Children Prescribed ADHD Medication* and *Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics* Measure Performance Scores – **Medicaid** Plans, calendar year 2018 data

Rate	Correlation Coefficient
	Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics
Follow-Up Care for Children Prescribed ADHD Medication	
Initiation	0.31*
Continuation and Maintenance	0.30*

*Significant at p<0.05

Results of face validity assessment

Input from our multi-stakeholder measurement advisory panels and those submitting to public comment indicate the measure has face validity.

2016 Submission

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)
2020 Submission

Interpretation of construct validity testing

Correlations between individual rates within the *Follow-Up Care for Children Prescribed ADHD Medication* measure were strong (Tables 4a, 4b) across product lines. In the commercial product line, correlations between the *Follow-Up Care for Children Prescribed ADHD Medication* and the *Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics* measure rates (Table 5a) were weak to moderate. In the Medicaid product line, correlations between the *Follow-Up Care for Children Prescribed ADHD Medication* and the *Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics* measure rates (Table 5b) were strong. Plans with higher rates on *Follow-Up Care for Children Prescribed ADHD Medication* tend to also have higher rates on the *Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics* measure. The results indicate that the *Follow-Up Care for Children Prescribed ADHD Medication* measure has good validity.

Interpretation of systematic assessment of face validity

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

2016 Submission

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.

2020 Submission

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

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

☐ **No risk adjustment or stratification**

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

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

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

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

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

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

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

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

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

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

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

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

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

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

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

2b3.9. Results of Risk Stratification Analysis:

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

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

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

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

2020 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.

2016 Submission

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

2020 Submission

Table 7. Calendar year 2018 Variation in Performance Across Health Plans

Product Line	Rate	Avg. EP	Avg. (%)	SD (%)	10th (%)	25th (%)	50th (%)	75th (%)	90th (%)	IQR (%)	p-value
Commercial	Initiation	417	40.0	8.3	29.8	35.0	40.0	44.3	50.4	15.4	0.0000
	Continuation & Maintenance	162	48.0	9.3	36.2	43.2	47.9	53.0	60.0	9.8	0.0017
Medicaid	Initiation	1,466	44.2	9.7	33.9	37.9	43.4	49.9	56.6	12.0	0.0000
	Continuation & Maintenance	340	54.6	12.0	39.0	46.4	55.5	62.7	71.2	16.3	0.0000

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.

2016 Submission

2016 Update: ABILITY TO IDENTIFY STATISTICALLY SIGNIFICANT/MEANINGFUL DIFFERENCES

HEDIS 2016 Variation in Performance across Health Plans- Commercial

Product Line	Rate	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Commercial	Initiation	397	39.0%	8.6%	29.1%	34.3%	38.6%	43.5%	50.2%	9.2%	<0.001
	C&M	163	46.8%	9.3%	35.6%	40.7%	46.4%	52.3%	57.3%	11.6%	0.002
Medicaid	Initiation	1,160	42.2%	11.0%	28.8%	34.2%	42.2%	49.6%	55.5%	15.4%	<0.001
	C&M	320	50.9%	13.3%	34.0%	40.9%	52.5%	62.5%	67.2%	21.6%	<0.001

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

Figure 1a. *Follow-up Care for Children Prescribed ADHD Medication - Initiation Phase*: Commercial Plans 2014-2016

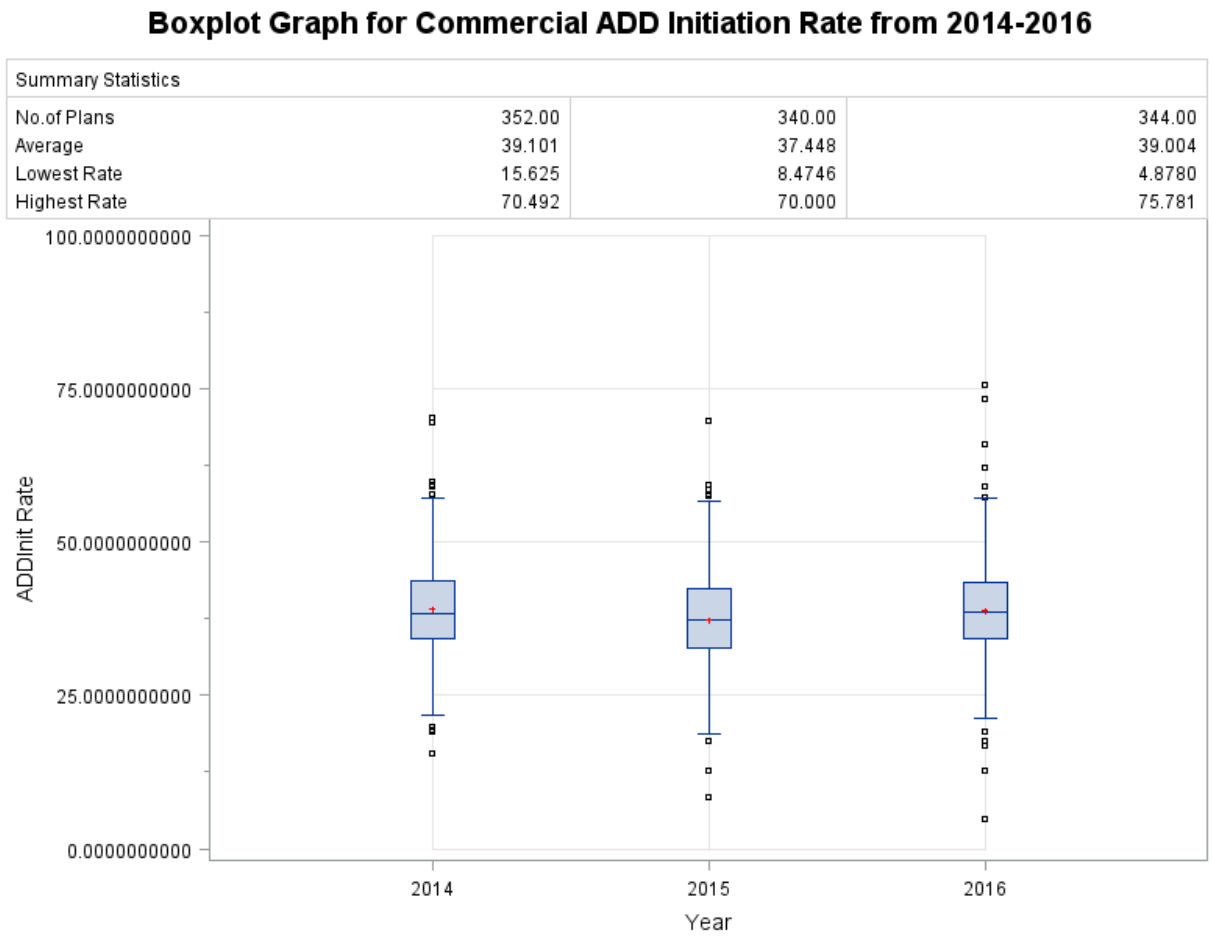


Figure is a graphical representation of ADD Initiation Rate by Year from 2014 to 2016.

Figure 1b. ***Follow-up Care for Children Prescribed ADHD Medication – C&M Phase*****: Commercial Plans 2014-2016**.

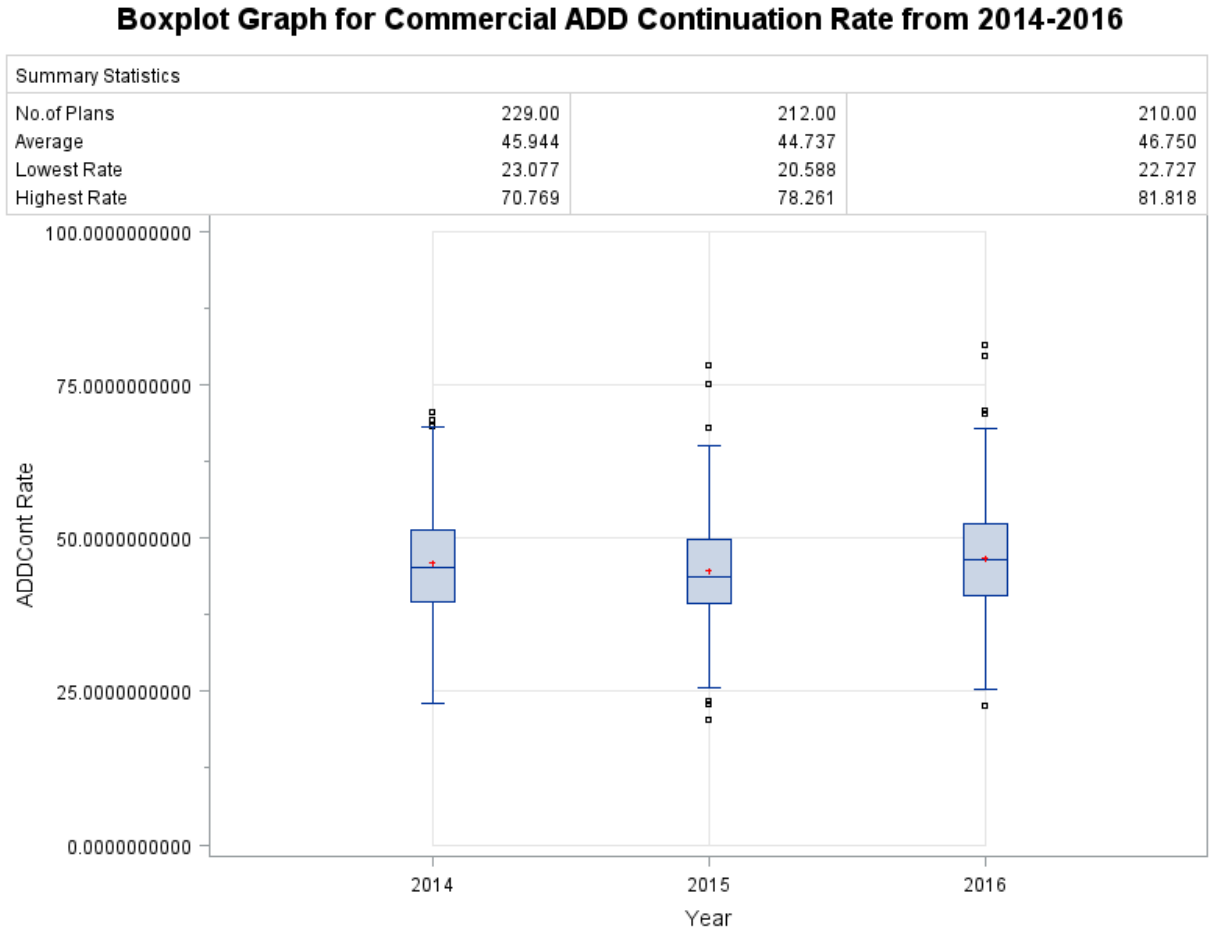


Figure is a graphical representation of Commercial ADD continuation rate by year from 2014 to 2016.

Figure 2a. ***Follow-up Care for Children Prescribed ADHD Medication - Initiation Phase*****: Medicaid Plans 2014-2016**

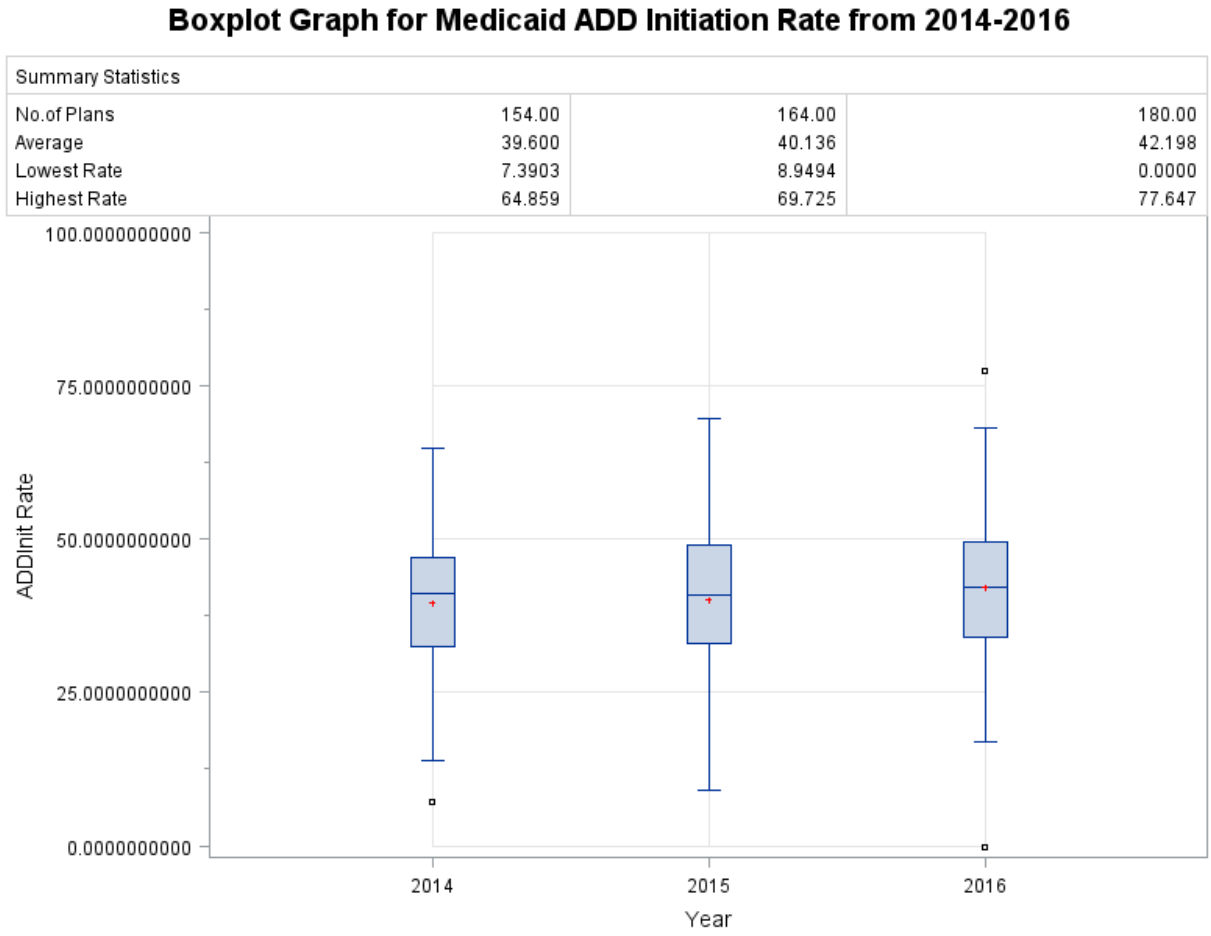


Figure is a graphical representation of Medicaid ADD initiation rate by year from 2014 to 2016.

Figure 2b. *Follow-up Care for Children Prescribed ADHD Medication – C&M Phase*: Medicaid Plans 2014-2016

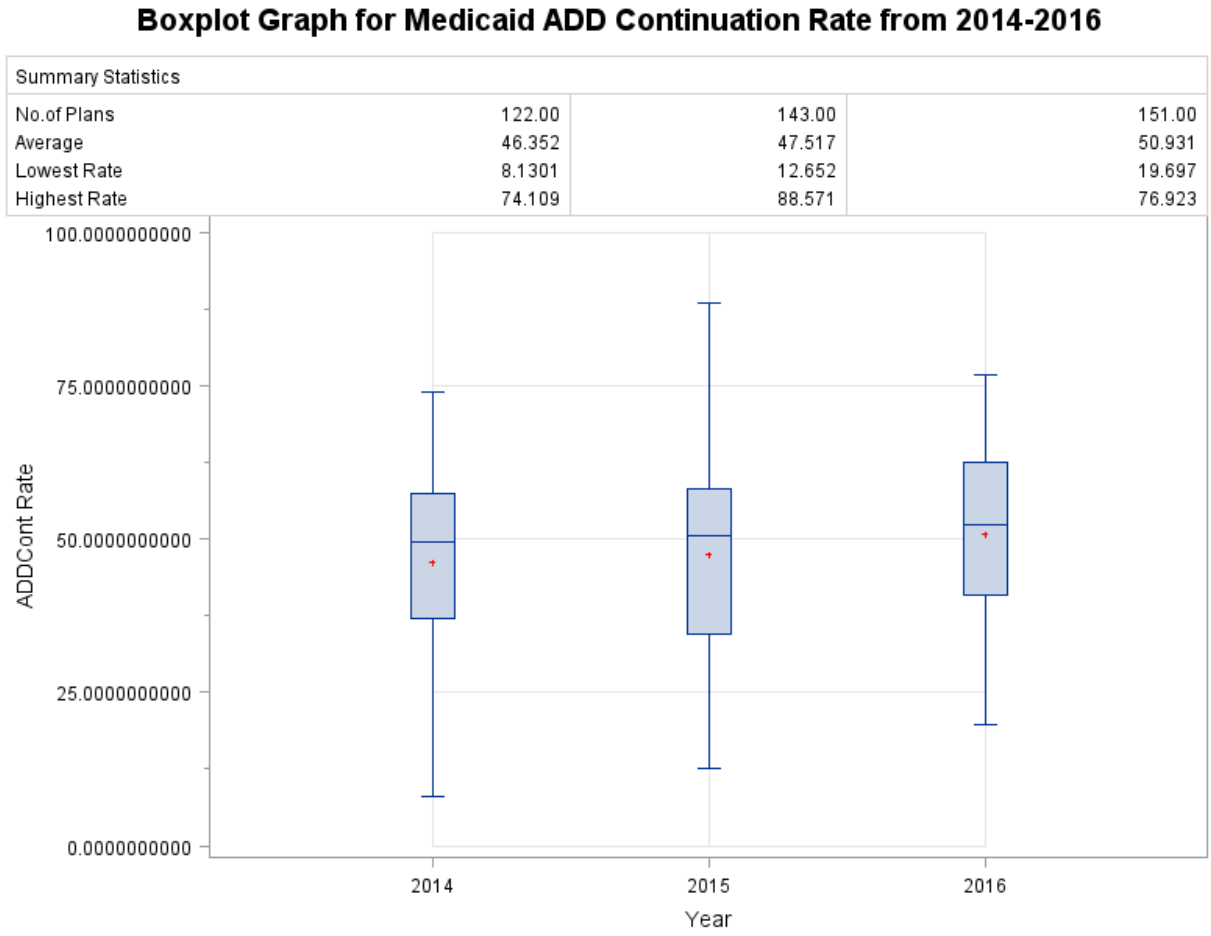


Figure is a graphical representation of Mediaic ADD continuation rate by year from 2014 to 2016.

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

2020 Submission

The results above indicate there is a 9.8-16.3% gap in performance between the 25th and 75th performing plans. For all product lines and rates, the difference between the 25th and 75th percentile is statistically significant. The largest gap in performance is for Medicaid plans rates, which show a 12-16.3 percentage point gap between 25th and 75th percentile plans. This gap represents on average 176 and 54 more children receiving Initiation Phase and Continuation and Maintenance Phase care respectively in high performing Medicaid plans compared to low performing plans

2016 Submission

2016 Update: INTERPRETATION OF ABILITY TO IDENTIFY STATISTICALLY SIGNIFICANT/MEANINGFUL DIFFERENCES

The results above indicate there is a 9-22% gap in performance between the 25th and 75th performing plans. For all product lines and rates the difference between the 25th and 75th percentile is statistically significant. The largest gap in performance is for the Medicaid health plans which show a 15.4-21.6 percentage point gap between 25th and 75th percentile plans. This gap represents on average 179 children in the Initiation Phase and

69 children in the Continuation and Maintenance Phase in high performing Medicaid plans compared to low performing plans (estimated from average health plan eligible population). Additionally, on average, plans in the 90th percentile performed approximately 21 percentage points better than plans in the 10th percentile in the commercial product line. In the Medicaid product line, on average, plans in the 90th percentile performed approximately 30 percentage points better than plans in the 10th percentile. Overall, these results suggest there are meaningful differences in performance and there is an opportunity 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*.**

2020 Submission

This measure has only one set of specifications.

2016 Submission

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

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

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

2020 Submission

HEDIS measures apply to enrolled members in a health plan, and NCQA has a rigorous audit process to ensure the eligible population and numerator events for each measure are correctly identified and reported. The

audit process is designed to verify primary data sources used to populate measures and ensure specifications are correctly implemented.

The HEDIS Compliance Audit addresses the following functions:

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

2016 Submission

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

2020 Submission

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

2016 Submission

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

2020 Submission

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

2016 Submission

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

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

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 organization´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 to questions in order to prevent possible errors or inconsistencies in the implementation of the measure. Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the usability and feasibility of the measure.

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

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

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 Health Plan Ranking http://reportcard.ncqa.org/plan/external/plansearch.aspx Health Plan Ranking http://reportcard.ncqa.org/plan/external/plansearch.aspx Payment Program CMS Quality Payment Program https://qpp.cms.gov/ CMS Promoting Interoperability Programs https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms Physician Value-Based Payment Modifier (VBM) https://www.cms.gov/medicare/medicare-fee-for-service-payment/physicianfeedbackprogram/valuebasedpaymentmodifier.html CMS Quality Payment Program https://qpp.cms.gov/ CMS Promoting Interoperability Programs https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms Physician Value-Based Payment Modifier (VBM) https://www.cms.gov/medicare/medicare-fee-for-service-payment/physicianfeedbackprogram/valuebasedpaymentmodifier.html Regulatory and Accreditation Programs Accreditation: http://www.ncqa.org/tabid/123/Default.aspx Accountable Care Organization Accreditation: http://www.ncqa.org/Programs/OtherPrograms/acomeasuresPilotProject.aspx Accreditation: http://www.ncqa.org/tabid/123/Default.aspx Accountable Care Organization Accreditation: http://www.ncqa.org/Programs/OtherPrograms/acomeasuresPilotProject.aspx Quality Improvement (external benchmarking to organizations) Quality Compass http://www.ncqa.org/tabid/177/Default.aspx Annual State of Health Care Quality http://www.ncqa.org/tabid/836/Default.aspx

4a1.1 For each CURRENT use, checked above (update for <u>maintenance of endorsement</u>), 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 QUALITY PAYMENT PROGRAM: This measure is used in the Quality Payment Program (QPP) which is a reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs).

CMS PROMOTING INTEROPERABILITY PROGRAM: This measure is used in the CMS Promoting Interoperability Program, which provides incentive payments to eligible professionals, eligible hospitals, and critical access hospitals (CAHs) as they adopt, implement, upgrade or demonstrate meaningful use of certified EHR technology, with an increased focus on interoperability and improving patient access to health information.
PHYSICIAN VALUE-BASED PAYMENT MODIFIER (VBM): This measure is used in the Physician Value-Based Modifier program, which provides differential payment to a physician or group of physicians under the Medicare Physician Fee Schedule (PFS). VBM is based on the quality of care provided in comparison to the cost of care within a performance period. The Value Modifier is an adjustment made to Medicare payments for items and services under the Medicare PFS.
HEALTH PLAN RATING/REPORT CARDS: This measure is used to calculate health plan rankings which are reported on the NCQA website. These rankings are based on performance on HEDIS measures among other factors. In 2019, a total of 255 Medicare health plans, 515 commercial health plans and 188 Medicaid health plans across 50 states were included in the rankings.
STATE OF HEALTH CARE ANNUAL REPORT: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2019, the report included results from calendar year 2018 for health plans covering a record 136 million people, or 43 percent of the U.S. population.
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 2012 the report included measures on 11.5 million Medicare Advantage beneficiaries in 455 Medicare Advantage health plans, 99.4 million members in 404 commercial health plans, and 14.3 million Medicaid beneficiaries in 136 plans across 50 states.
MEDICAID CHILD CORE SET: This measure is included in the Medicaid Child Core Set which is a set of children's health care quality measures developed as part of the Children's Health Insurance Program (CHIP) Reauthorization Act for voluntary use by State Medicaid and CHIP programs. The data collected with these measures will help CMS to better understand the quality of health care children receive through Medicaid and CHIP and assist CMS and states in moving toward a national system for quality measurement, reporting, and improvement. As per the CHIPRA legislation, state data derived from the core measures will become part of the Secretary's annual report on the quality of care for children in Medicaid and CHIP. The Secretary's annual report summarizes state-specific and national measurement information on the quality of health care furnished to children enrolled in Medicaid and CHIP.
NCQA HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. In 2012, a total of 170 Medicare Advantage health plans were accredited using this measure among others covering 7.1 million Medicare beneficiaries. [REPLACE or ADD as appropriate, 336 commercial health plans covering 87 million lives; 77 Medicaid health plans covering 9.1 million lives.] Health plans are scored based on performance compared to benchmarks.
NCQA ACCOUNTABLE CARE ORGANIZATION ACCREDITATION: This measure is used in NCQA's ACO Accreditation program, that helps health care organizations demonstrate their ability to improve quality, reduce costs and coordinate patient care. ACO standards and guidelines incorporate whole-person care coordination throughout the health care system.
QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer simple comparison of plans' performance against competitors or benchmarks.
PHYSICIAN FEEDBACK/QUALITY AND RESOURCE USE REPORTS (QRUR): This measure is used in the Physician Feedback Program and Quality and Resource Use Reports which provide comparative performance information to

Medicare Fee-For-Service physicians. The Quality and Resource Use Reports show physicians the portion of their Medicare fee-for-service (FFS) patients who have received indicated clinical services, how patients utilized services, and how Medicare spending for their patients compares to average Medicare spending. QUALIFIED HEALTH PLAN (QHP) QUALITY RATING SYSTEM (QRS): This measure is used in the Qualified Health Plan (QHP) Quality Rating System, which provides comparable information to consumers about the quality of health care services and QHP enrollee experience offered in the Marketplaces.

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

N/A

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

N/A

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

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

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

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

NCQA publishes HEDIS results annually in our Quality Compass tool. NCQA also presents data at various conferences and webinars. For example, at the annual HEDIS Update and Best Practices Conference, NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly. NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System, as described in Section 3c1.

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. During this "reevaluation" process, we seek broad input on the measure, including input on performance and implementation experience. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System. This information enables NCQA to comprehensively assess a measure's adherence to the HEDIS Desirable Attributes of Relevance, Scientific Soundness and Feasibility.

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

In general, health plans have not reported significant barriers to implementing this measure, as it uses the administrative data collection method. Questions have generally centered around minor clarification of the specifications, such as confirmation that information in claims meets the measure intent and questions about the supporting guidelines for the measure. NCQA responded to all questions to ensure consistent implementation of the specifications.

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

We have provided minor clarifications about the measure during the annual update process in order to address questions received through the Policy Clarification Support System.

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.

Feedback has not required modification to this 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.

2017 to 2019 data shows relatively stable performance and room for improvement across Commercial and Medicaid plans. For the 3 years of data analyzed, the mean performance for rate 1 – Initiation Phase was 40% for Commercial plans and 44.4% for Medicaid plans, while the mean performance for rate 2 – Continuation and Maintenance Phase was 47.1% and 54.7% for Commercial and Medicaid plans respectively. Performance rates vary slightly from year to year; however, Commercial plans reported the highest increase in performance for having at least two follow-up visits on different dates of service with any practitioner from 31–300 days (9 months) after the earliest prescription dispensing date for a new ADHD medication. Across both commercial and Medicaid plans, there continues to be fairly large variation between the 10th and 90th percentiles, suggesting room for improvement. For example, among commercial plans, the 2019 rate of children who had documentation of a timely follow-up visit ranged from 20% for plans in the 10th percentile to 60% among plans in the 90th percentile.

4b2. Unintended Consequences

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

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

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

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

There were no identified unexpected findings during testing or since implementation of this measure.

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?

No

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: Brittany, Wade, wade@ncqa.org, 202-530-0463-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

Describe the members´ role in measure development.

NCQA BEHAVIORAL HEALTH MEASUREMENT ADVISORY PANEL

Katharine Bradley, MD, MPH, Kaiser Permanente Washington Health Research Institute

Christopher Dennis, MD, MBA, FAPA, Landmark Health

Ben Druss, MD, MPH, Emory University

Frank A. Ghinassi, PhD, ABPP, Rutgers University Behavioral Health Care

Connie Horgan, ScD, Brandeis University

Laura Jacobus-Kantor, PhD, SAMSA, HHS

Jeffry Meyerhof, MD, Optum Behavioral Health

Harold Pincus, MD, Columbia University

Michael Schoenbaum, PhD, National Institute of Mental Health

John H. Straus, MD, Beacon Health Options

COMMITTEE ON PERFORMANCE MEASUREMENT

Andrew Baskin, MD, Aetna

Andrea Gelzer, MD, MS, FACP, AmeriHealth Caritas

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

David Grossman, MD, MPH, Washington Permanente Medical Group

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

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

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

Nancy Lane, PhD, Independent Consultant

Bernadette Loftus, MD, The Permanente Medical Group

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

Amanda Parsons, MD, MBA, Montefiore Medical Center

Wayne Rawlins, MD, MBA, ConnectiCare

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

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

JoAnn Volk, MA, Georgetown University

Lina Walker, PhD, AARP

NCQA TECHNICAL MEASUREMENT ADVISORY PANEL

Andy Amster, MSPH, Kaiser Permanente

Jennifer Brudnicki, MBA, Inovalon Inc.

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

Kathryn Coltin, MPH, Independent Consultant

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

Marissa Finn, MBA, CIGNA

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

Carlos Hernandez, CenCal Health

Harmon Jordan, ScD, Westat

Virginia Raney, LCSW, Center for Medicaid and CHIP Services

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

Laurie Spoll, Aetna

HEDIS EXPERT CODING PANEL

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

Denene Harper, RHIA, American Hospital Association

DeHandro Hayden, BS, American Medical Association

Patience Hoag, RHIT, CPHQ, CHCA, CCS, CCS-P, Aqurate Health Data Management

Nelly Leon-Chisen, RHIA, American Hospital Association

Alec McLure, RHIA, CCS-P, Verisk Health

Michele Mouradian, RN, BSN, McKesson Health Solutions

Craig Thacker, RN, CIGNA HealthCare

Mary Jane F. Toomey, RN CPC, WellCare Health Plans, Inc.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2006

Ad.3 Month and Year of most recent revision: 07, 2014

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

Ad.5 When is the next scheduled review/update for this measure? 12, 2017

Ad.6 Copyright statement: ©2006 by the National Committee for Quality Assurance

1100 13th Street, NW, Suite 1000

Washington, DC 20005

Ad.7 Disclaimers: These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.

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

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

These performance measures were developed and are owned by NCQA. They are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and NCQA has no liability to anyone who relies on such measures. NCQA holds a copyright in these measures and can rescind or alter these measures at any time. Users of the measures shall not have the right to alter, enhance or otherwise modify the measures, and shall not disassemble, recompile or reverse engineer the source code or object code