NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 2800

Corresponding Measures:

De.2. Measure Title: Metabolic Monitoring for Children and Adolescents on Antipsychotics

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

De.3. Brief Description of Measure: The percentage of children and adolescents 1-17 years of age who had two or more antipsychotic prescriptions and had metabolic testing.

1b.1. Developer Rationale: This measure addresses metabolic monitoring as one facet of safe and judicious use of antipsychotics in children and adolescents. Although antipsychotic medications offer the potential for effective treatment of psychiatric disorders in children, they can also increase a child's risk for developing serious metabolic health complications associated with poor cardiometabolic outcomes in adulthood. Despite the risk of such adverse side effects and clinical guideline recommendation, evidence suggests that children and adolescents do not receive appropriate laboratory monitoring. Thus, this measure encourages metabolic monitoring of children who are on antipsychotic medications.

S.4. Numerator Statement: Children and adolescents 1-17 years of age on antipsychotics who received blood glucose and cholesterol testing during the measurement year.

S.6. Denominator Statement: Children and adolescents 1-17 years of age who had ongoing use of antipsychotic medications (at least two prescriptions).

S.8. Denominator Exclusions: Patients in hospice.

De.1. Measure Type: Process

S.17. Data Source: Claims

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: May 04, 2016 **Most Recent Endorsement Date:** May 04, 2016

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

- Previously reviewed by the Pediatrics Standing Committee.
- The developer provided the following logic model: Child or adolescent has ongoing use of antipsychotic medication >>> Metabolic monitoring by a health care provider >>> Identification of metabolic issues/side effects >>> Health care provider addresses metabolic issue >>> Patient receives intervention for metabolic issues present >>> Metabolic issues reduced or eliminated >>> Improvement in metabolic functioning for patient.
- The measure is based on three evidence-based clinical practice guideline recommendations from the American Academy of Child and Adolescent Psychiatry (AACAP).
- In 2016, the Committee agreed this is an important measure to monitor the serious side effects of prescribing antipsychotic medications to children and adolescents (e.g., diabetes, rapid weight gain) and agreed evidence exists to support glucose and lipid monitoring for children on

antipsychotics. Clear recommendations are provided by the professional societies regarding concern for metabolic derangements.

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: N/A

Exception to evidence

N/A

Questions for the Committee:

- The developer attests the evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need to repeat the vote on Evidence?
- Numerous guideline recommendations based on a large body of evidence support metabolic monitoring, and there is a strong conceptual model. However, direct evidence of the improved outcomes is not specifically stated. Is the link between this process and patient outcomes clear and strong?

Guidance from the Evidence Algorithm

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

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

RATIONALE:

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.

- HEDIS data presented for 2016-2018 reflects that for commercial and Medicaid plan mean performance is 33-35% each year with a standard deviation of 10-13%.
- Performance data is similar across years presented.
- It would be useful to see performance rates by age range and by test type as the measure is specified to be reported by these variables.

Disparities

- Based on HEDIS field testing Hispanic children had better rates of baseline metabolic screening (10.3 percent) compared to white non-Hispanic children (5.7 percent) and black non-Hispanic children (6.1 percent) and better rates of ongoing metabolic monitoring (24.8 percent)

compared to white non-Hispanic children (19.1 percent) and black non-Hispanic children (19.4 percent). Another study cited found that race was not associated with screening rates.

- A 2011 study found children with Medicaid in foster care received screening at a higher rate (28%) than children with Medicaid not in foster care (18%).
- The developer notes that HEDIS data are stratified by insurance type (Medicaid and Commercial), which serves as a proxy for socioeconomic status. Performance rates from 2016-2018 are similar for both insurance types.

Questions for the Committee:

- Overall performance is low with variation across percentiles, which reflects opportunity for improvement. Are plans able to improve performance?
- Is the Committee satisfied that the data provided indicates an overall gap in care without specific information by age or test type?

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

RATIONALE:

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."

**seems to be good evidence. I assume all types of antipsychotics (e.g., clozapine) have metabolic side effects.

**This is a process measure and the evidence gathered relates to the process being measured. This process measure is important as it allows identification of adverse effects of antipsychotic medication for children and teens so that the ultimate outcome can be improvement in metabolic functioning for the patients.

**maintenance measure; no change to evidence.

**"The developer's attestation re: no changes in the evidence. The main source for evidence in AACAP practice parameters, July 2012, including 147 publications. The developer reports that the AACAP practice parameter does not include any studies that examine the positive or negative (i.e., harm) for metabolic monitoring. New studies that were not included in submission: Nov 2019 study in JAACAP, led by Karsten Jensen: 113 youth with first episode psychosis, quetiapine-ER vs. aripiprazole. 12-week (short term) trial. Quetiapine ER: Significantly greater increases in weight Total cholesterol, triglycerides, LDL (not HDL) No difference in glucose at 12 weeks Increase in hyperinsulinemia--homeostatic model assessment of insulin resistance (HOMA-IR) "Because more valid marker if risk of developing diabetes and metabolic syndrome". 2015 Rettew, et al. Pediatrics,

prescriber survey (80% response rate), Vermont Medicaid only. Provider reported metabolic monitoring: 57.2% 2014 Delate, et al. JAMA Peds, commercially insured children/Kaiser Colorado: 2002-2011 At least one indicator: fasting triglyceride, fasting blood glucose, BP or wt (cannot assess which one or if wt and BP simply part of routine VS): Baseline: 2002: 50.5%; 2011: 60.5% Follow-up (84 days after index/baseline monitoring): 2002: 45.1%; 2011: 55.3%"

**This is a process maintenance measure based on claims data from the health plan level. The numerator is children and adolescents age 1-17 on antipsychotics who received blood glucose and cholesterol testing during the measurement year while the denominator is children and adolescents age 1-17 who had ongoing use of antipsychotic medication during the measurement year. The original endorsement of this measure was in May, 2016. At that time, there was a graded systematic review of 3 evidence-based clinical practice guidelines from the American Academy of Child and Adolescent Psychiatry. There was strong empirical evidence relating directly to the process being measured. There is also evidence that it relates to the desired outcome of improving health care. The developer reported that there are no new studies or information that changes the evidence base.

**Not aware of changes in evidence base for this measure.

**Direct evidence of the improved outcomes from metabolic monitoring is not specifically stated.

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?

**yes

**Yes and there is variance in data collection on this measure.

**gap demonstrated.

**Meaningful differences in Performance NCQA: What is the benchmark increase for being "effective for promoting improvement"? Vs. if little change, does this support persistent performance gap? 4 years of data (2015-2018) Medicaid: 2015: 29.8%; 2016: 33.3%; 2017: 34.6%; 2018: 35.3% Commercial HMO: 2015: 33.9%; 2016: 36%, 2017: 37%; 2018: 37.2% Commercial PPO: 2015: 30.7%, 2016: 32.3%; 2017: 32.8%; 2018: 33.1% NCQA argues there are meaningful difference when comparing the 25th and 75th percentile f or the measures in 2018. Very large sample size, P's are <.0001, but are of course influenced by sample size Disparities not assessed because NCQA data source for HEDIS measures limited to health plan level

**Performance data was provided based on HEDIS data from 2016 and 2018. For both commercial plans and Medicaid the rates were 33-35% showing a gap and an opportunity for improvement. As for disparities, Hispanic youth had better rates (10.3%) than Whites (5.7%) and African Americans (6.1%). Ongoing rates for Hispanics were also better (24.8%) than Whites (19.1%) and African Americans (19.4%). In Medicaid, foster youth had higher rates (28%) than non-foster youth (18%).

**Data by population subgroups provided, low level of performance overall.

**Overall performance is low which reflects opportunity for improvement

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.

Composite measures only:

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

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

Evaluators: NQF Staff

NQF Staff Review

Evaluation Summary:

Reliability

- The developer provided updated reliability testing.
- Score level reliability results from signal-to-noise testing in 273 commercial plans and 169 Medicaid plans indicate that the measure has good reliability in both populations. Overall reliability statistics were 0.875 for commercial plans and 0.985 for Medicaid plans. The lower percentiles for the commercial population slightly lower scores.

Validity

- The developer provided updated validity testing.
- Score level results of construct validity testing indicate moderate-to-weak, significant correlations between the measure and other measures with preventive care components:

Adolescent Well-Care Visits and Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life.

- Face validity testing previously provided was supportive of the measure, but voting results were not provided.
- The IQR shows significant performance differences for both commercial and Medicaid plans.

Questions for the Committee regarding reliability:

- Do you have any concerns about the measure specifications?
- Is there any concern about the average reliability statistics (i.e., 0.54, 0.64) for the lower percentiles (i.e., 10th and 25th) of the commercial population?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure based on the testing provided?
- Is the hospice exclusion and stratification of results appropriate?

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

Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 2800

Measure Title: Metabolic Monitoring for Children and Adolescents on Antipsychotics

- Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use** ☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

- Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data** ☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data** ☐ **Enrollment Data** ☐ **Other**

- Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☐ **Facility** ☒ **Health Plan** ☐ **Population: Community, County or City** ☐ **Population: Regional and State** ☐ **Integrated Delivery System** ☐ **Other**

- Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

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

RELIABILITY: TESTING

- Is the measure specified to be reported by age (1-11 years, 12-17 years, and total), test type (blood glucose, cholesterol, and total), and payer (commercial and Medicaid)?

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

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

   ☐ **Yes** ☐ **No** N/A

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

   Submission document: Testing attachment, section 2a2.2

   - Measure score reliability was tested using 2018 HEDIS data that included 273 commercial plans (mean denominator size/plan was 245) and 169 Medicaid plans (mean denominator size/plan was 1128).
   - Reliability of the measure score was tested using a beta-binomial calculation. The reliability is represented as a signal to noise ratio.
   - The method used is appropriate. The testing matches the level of analysis (health plan) and uses a large data set.

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

   Submission document: Testing attachment, section 2a2.3

   Overall Beta-binomial statistic and distribution of plan reliability for commercial and Medicaid product lines, 2018

Product Line	Overall Reliability	Min	Percentiles					Max
			10th	25th	50th	75th	90th	
Commercial	0.875	0.439	0.540	0.644	0.795	0.892	0.948	0.985
Medicaid	0.985	0.648	0.864	0.936	0.975	0.990	0.994	0.998

   - Results indicate that both product lines have good reliability. The lower percentiles for the commercial population have lower reliability scores.

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.

    - Measure excludes patients in hospice. Developer does not provide additional information regarding frequency or variation in the testing form, but this seems appropriate.

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

    Submission document: Testing attachment, section 2b4.

    - No concerns.
    - Results indicate that there is significant 13.5% gap in performance between Medicaid plans at the 25th and 75th percentiles and a significant 12.2% gap in performance among commercial plans.

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.

    N/A

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

    Submission document: Testing attachment, section 2b6.

    - Developer describes audit process to identify bias or errors, but they do not share results or additional information about missing data.
    - Is it possible tests are being completed but not billed appropriately to be included or that there are errors due to omission?

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**

- The developer states the measure is not risk adjusted, but it is also specified to be reported by type of insurance, which serves as a proxy for socioeconomic status.

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

Empirical Validity

- The developer tested construct validity of the measure by exploring whether it was correlated with the following measures: 1) Adolescent Well-Care Visits and 2) Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life.
- Hypothesis: Plans that perform well on the measure should perform well on these other measures of similar construct (routine care/monitor for youth).
- A Pearson correlation test was performed to test the strength of the relationship.
- This testing approach is appropriate based on NQF standards.

Face Validity (2015)

- NCQA and five expert panels worked together in 2013-2014 to identify the most appropriate method for assessing metabolic monitoring among this patient population. The measure also went through public comment in 2014.

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

Submission document: Testing attachment, section 2b2.3

Empirical Validity

Results of Pearson Correlation Coefficient for commercial health plans, 2018.

Measure	Pearson Correlation Coefficients	
	Adolescent Well-Care Visits	Well-Child Visit
Metabolic Monitoring	0.41	0.37

**includes data submitted by 273 Commercial plans to HEDIS for these measures for measurement year 2018*
Note: all correlations are significant at p<.0001

Results of Pearson Correlation Coefficient for Medicaid health plans, 2018.

Measure	Pearson Correlation Coefficients	
	Adolescent Well-Care Visits	Well-Child Visit
Metabolic Monitoring	0.31	0.39

**includes data submitted by 169 Commercial plans to HEDIS for these measures for measurement year 2018*
Note: all correlations are significant at p<.0001

- Results indicate moderate-to-weak, significant correlations between the measure and comparators and confirm the developer's hypothesis.

Face Validity (2015)

- Across the multiple expert panels that reviewed this measure, all panels concluded this measure was specified to assess metabolic monitoring. Specific voting results were not provided, but the developer indicates that the expert panels showed good agreement that the measure as specified will accurately differentiate quality across health plans. Public commenters supported the measure.

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.**

ADDITIONAL RECOMMENDATIONS

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

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/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

**were the specifications updated to take account of new antipsychotic medications or formulations that entered the market? Is glucose and lipid testing always separately identifiable on claims data and captured in claims data?

**This measure has good reliability and there are no concerns that it can be consistently implemented.

**no concerns

**"Reliability: prior testing: 2008 MAX data (11 states); 2010 Medicaid health plan (17 plans), commercial health plan (73 plans) 2012 For 2019 submission: Signal-to-noise testing (beta-binomial calculation), included 2018 HEDIS data, assessed proportion of total variation attributable to a health plan "signal", commercial: .875 Medicaid: .985, similar to 2015 submission findings. Half the commercial health plan and neatly all Medicaid health plans exceed the 0.7 cut-point for very good reliability But let's look at the details: There are actually 3 numerators that differentiate types and extent of metabolic testing: At least one test for blood glucose/any time during the measurement year At least one test for LCL or cholesterol/measurement year Both tests/any time during the measurement year Clinically this is a very low bar, and a child could pass if only a baseline glc was drawn? This would not monitor change in metabolic parameters while on ap medication. Which numerator is NCQA publicly reporting? Denominator: at least 2 ap med dispensing events of same or different ap medication on different dates during measurement year, no more than one 45 day gap

in enrollment. Clinically: ap dispense is not med adherence, the time point of whether baseline or follow-up lab relative to dispense date is not assessed. So you can't tell if really monitoring change while on a medication."

**No concerns

**Well defined, no concerns.

**Submitted specifications are precise

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

**no

**No

**no concerns

**please clarify which numerator is used when reporting adherence rate

**No

**No concerns.

**New reliability results provided. Overall reliability statistics were 0.875 for commercial plans and 0.985 for Medicaid plans.

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

**the face validity is strong. the empirical testing was weak but not necessary given the strong face validity

**The measure has face and construct validity.

**no concerns

**does not address the question, "If there is adherence to this measure, is there evidence that there is improved medication safety monitoring for children receiving on-going ap medication during an episode of care?

**No

**No concerns.

**Results indicate moderate-to-weak. The developer tested construct validity of the measure by exploring whether it was correlated with the following measures: 1) Adolescent Well-Care Visits and 2) Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life. Question-- is it possible that well-visits occur with appropriate frequency but what occurs during the well visit is not well correlated i.e.metabolic screening

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?

**no concerns

**I do not believe there are major threats to validity of this measure.

**no concerns

**NCQA argues there are meaningful difference when comparing the 25th and 75th percentile for the measures in 2018. Very large sample size, P's are <.0001, but are of course influenced by sample size Missing data: NCQA argues rigorous audit process for HEDIS measures

**As for threats to validity, the developer described the audit process to identify bias or errors but did not share results. For validity testing, the expert panels were supportive of the measure but voting results were not provided. Despite this, they do not constitute a threat to validity.

**no concerns.

**Possible missing data due to tests being performed but not billed.

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 risk-adjustment 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?

** risk adjustment is not needed

**Social risk factors were not measured.

**no concerns

**appropriately excludes hospice, no capacity to adjust by social risk factors given NCQA data source

**Exclusions (patients in hospice) were appropriate. There was no risk adjustment.

**No concerns.

**NA

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.

- Data elements are generated during the provision of care and are defined fields in electronic claims.
- The measure has been widely implemented.
- NCQA has an audit process to verify data. Results for this particular measure were not provided.

Questions for the Committee:

- Are the data elements required always captured in claims?

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

RATIONALE:

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?

**no concerns. already in use. Although see question about whether claims always capture glucose and lipid tests

**NCQA conducts an independent audit of all HEDIS collection and reporting processes. I don't have any concerns about operational use.

**no concerns

**Use by NCQA reports. Publicly reported Only at the health plan level, not patient level using claims data?

**No concerns. The routine data elements can be routinely generated.

**Measure has been widely implemented, no concerns.

**The measure has been widely implemented

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

- Publicly reported in the NCQA State of Health Care annual report.
- Used to calculate health plan ratings, which are reported in Consumer Reports and on NCQA's website.
- Used in Quality Compass, a tool used for selecting a health plan, conducting competitor analysis, examining quality improvement, and benchmarking plan performance.

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

- Those being measured (i.e., health plans) have access to performance results and benchmarks allow plans to understand performance compared to other plans.
- The developer reports users have the opportunity to provide feedback. Feedback has consisted of minor clarifications, which have been explained during the annual update process.

***Questions for the Committee**:*

- Are the measure results transparent and can they assist in driving improvements in care for children on antipsychotic medication?

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

RATIONALE:

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

- Nationally representative data shows slight improvement over time (mean increase of 1-2% from 2016 to 2018 for both commercial and Medicaid plans).

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

Unexpected findings (positive or negative) during implementation

- The developer does not report unexpected findings.

Potential harms

- The developer does not report potential harms.

Additional Feedback:

- The 2018 MAP recommended this measure for addition to the Medicaid Child Core Set, noting it is both feasible and important. This measure will be added to the Medicaid Child Core Set in 2020.
- Public comments supported the adoption of this measure because it highlights the impact of comorbidities associated with children on antipsychotic medications. Two commenters suggested that this measure also be specified for those with a single medication prescription.

***Questions for the Committee**:*

- Are there potential unintended consequences of this measure (e.g., potential harms of switching antipsychotics)? Do the benefits outweigh the risks?
- Can health plans improve upon this measure?

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

RATIONALE:

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

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?

**Can NCQA provide more information on the feedback that it has received from health plans as it implemented this measure?

**The measure is publicly reported nationally and by geographic region.

**currently in use; will be added to 2020 core set

**reported publicly by NCQA

**The measure is reported by NCQA in its State of Health Care annual report. Data is used to calculate health plan ratings reported in Consumer Reports and the NCQA website. It is also used in Quality Compass. Feedback is given regarding performance results, opportunities to provide feedback which is considered in any changes incorporated into the measure.

**No concerns.

**Publicly reported and used in current accountability program

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.

**Seems usable

**There is good benefit to measurement. I do not see any concerning harms.

**slight improvement noted 1-2% 2016 - 2018

**Only slight improvement (mean increase 1-2% 2016-2018 commercial and Medicaid health plans) Note: the AACAP practice parameters do not provide evidence that monitoring is associated with positive outcomes (or harm)—simply not tested

**Nationally representative data shows a slight improvement over time from 2016 to 2018 with a mean increase of 1.2%. There are no unintended consequences.

**Results can be used to help assure metabolic testing on children and adolescents on antipsychotics.

**The link between metabolic testing and improvement in health outcomes not discussed.

Criterion 5: Related and Competing Measures

Related or competing measures

Related – 1932: Diabetes Screening for People With Schizophrenia or Bipolar Disorder Who Are Using Antipsychotic Medications (SSD)

Harmonization

The measures have different target populations (adults w/ schizophrenia or bipolar disorder prescribed antipsychotics versus adolescents prescribed antipsychotics) and focus on similar, but slightly different aspects of care (diabetes screening based on receipt of glucose test versus glucose test plus cholesterol test). The glucose test element is harmonized across the measures.

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?

**not that I know of

**Measure 1932: Diabetes Screening for People with Schizophrenia or Bi-Polar Disorder targets adults 18 to 64; the target population for measure 2800 is children and adolescents.

**1932 - diabetes screening for people with schizophrenia or bipolar using antipsychotic meds. (doesn't include cholesterol testing).

**There are related measures.

**Related measure - 1932, no competing measures.

**no concerns

Public and Member Comments

No NQF members have submitted support/non-support choices or public comments as of 01/23/2020.

Developer Submission

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

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

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

APM_Evidence_Form_-2800-.docx

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

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

No

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2800

Measure Title: Metabolic Monitoring for Children and Adolescents on Antipsychotics

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

Date of Submission: 8/1/2019

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: Annual blood glucose and cholesterol testing for children and adolescents on antipsychotics

☐ 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.

This measure assesses metabolic monitoring (i.e., the receipt of glucose and cholesterol tests) among children and adolescents that have ongoing antipsychotic use. Given the documented metabolic risks of antipsychotic medications, monitoring of metabolic indices is important to ensure appropriate management of side effect risk, especially in youth. The path envisioned is as follows.

Child or adolescent has ongoing use of antipsychotic medication >>> Metabolic monitoring by a health care provider >>> Identification of metabolic issues/side effects >>> Health care provider addresses metabolic issue >>> Patient receives intervention for metabolic issues present >>> Metabolic issues reduced or eliminated >>> Improvement in metabolic functioning for patient (desired outcome).

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

N/A

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

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

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 Child and Adolescent Psychiatry – Atypical Antipsychotic Medications 2011 Guideline on Fasting Profiles.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Practice Parameter for the Use of Atypical Antipsychotic Medications in Children and Adolescents** • **American Academy of Child and Adolescent Psychiatry** • **July 2012** • **N/A** • **https://www.aacap.org/App_Themes/AACAP/docs/practice_parameters/Atypical_Antipsychotic_Medications_Web.pdf**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Table: Fasting plasma glucose – Baseline, 12 wks, annually; Fasting lipid profile – Baseline, 12 wks (Recommendation 10, and Table 2): • "The acute and long-term safety of these medications in children and adolescents has not been fully evaluated and therefore careful and frequent monitoring of side effects should be performed…*Ideally, monitoring of BMI, blood pressure, fasting glucose and fasting lipid profiles should follow, whenever feasible, the recommendations found in the consensus statement put forth by the American Diabetes Association and American Psychiatric Association*."
Grade assigned to the **evidence** associated with the recommendation with definition of the grade	Grade of Recommendation: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time)
Provide all other grades and definitions from the evidence grading system	Grade of Recommendation: Minimal Standard/Clinical Standard: • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time Options • Acceptable but not requires; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated.

	AACAP Strength of Empirical Evidence: AACAP rates the strength of the empirical evidence in descending order as follows: • (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 non-randomly 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 the definition of the grade	Grade of Recommendation: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time)
Provide all other grades and definitions from the recommendation grading system	Grade of Recommendation: Minimal Standard/Clinical Standard • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time Options • Acceptable but not required; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated.
What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?	This measure addresses metabolic monitoring as one facet of safe and judicious use of antipsychotics in children and adolescents. Given the documented metabolic risks of antipsychotic medications, monitoring of metabolic indices is important to ensure appropriate management of side effect risk, especially in youth. Numerous guidelines address the need for metabolic monitoring among youth on antipsychotic medications. This measure is based on this American Academy of Child and Adolescent Psychiatry (AACAP) guideline, as well as other AACAP

	guidelines and by other organizations, such as the Canadian Alliance for Monitoring Effectiveness and Safety of Antipsychotics in Children (CAMESA). These organizations recommend metabolic testing for youth prescribed antipsychotics, with consensus that baseline and ongoing metabolic monitoring are standards of care for this population.
What is the time period covered by the body of evidence?	1990-2010.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Quantity: • When developing their guidelines, AACAP limited its evidence review to clinical trials, meta-analysis, practice guidelines, randomized controlled trials (RCTs), systematic literature reviews, and case reports and series. AACAP selected a total of 147 publications for careful examination based on their weight in the hierarchy of evidence attending to the quality of individual studies, relevance to clinical practice and the strength of the entire body of evidence. However, AACAP did not provide a breakdown of specific numbers of each publication type. Given the number of studies selected we did not feel comfortable re-conducting the evidence review and delineating all the publication types for each guideline. Instead we have identified where there are certain publication types available to support each guideline. • This recommendation is based on expert opinion established during a consensus development conference for four medical professional societies. The four societies found that an increasing number of methodologically rigorous studies have assessed the effectiveness of antipsychotics for children and adolescents in specific clinical situations. However, the long-term safety profile of each antipsychotic used by youth has yet to be effectively evaluated and characterized. In the absence of such evidence, AACAP recommends increased vigilance. • American Diabetes Association, American Psychiatric Association, American Association of Clinical Endocrinologists, North American Association for the Study of Obesity. Consensus development conference on antipsychotic drugs and obesity and diabetes. *Diabetes Care.* 2004;27:596-601. Quality: • The evidence review used by AACAP prioritized study designs less subject to bias and studies that represent the best scientific evidence. The evidence review included a large number of

	studies with large numbers of patients from various populations. Overall, the quality of the evidence regarding metabolic monitoring for children and adolescents on antipsychotics is high. The evidence provides a strong link between antipsychotic use and adverse metabolic side effects in youth and to negative long-term health outcomes throughout the lifespan.
Estimates of benefit and consistency across studies	AACAP did not cite any studies that directly evaluated the benefit of metabolic monitoring for children and adolescents on antipsychotics. However, the evidence demonstrates the adverse side effects, including diabetes, weight gain, and hyperlipidemia, as well as the concerns regarding the safety of long-term antipsychotics use in youth. Thus, AACAP estimates there is a greater benefit to be gained through increased vigilance and regular metabolic monitoring.
What harms were identified?	AACAP did not cite any studies that directly evaluated the harm of metabolic monitoring for children and adolescents on antipsychotics. AACAP noted that some patients and their parents may face negative social consequences due to frequent medical appointments, including greater time constraints on school and work responsibilities. However, given the adverse side effects and concerns regarding the safety of long-term antipsychotics use in youth, regular metabolic monitoring is still vital in the follow-up of these patients. Thus, in the absence of evidence of other harms, AACAP estimated there is less harm through increased vigilance and regular metabolic monitoring.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	To our knowledge, there have been no new studies that contradict the current body of evidence.

Table 2. American Academy of Child and Adolescent Psychiatry – Atypical Antipsychotic Medications 2011 Guideline on Blood Glucose Monitoring.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Practice Parameter for the Use of Atypical Antipsychotic Medications in Children and Adolescents** • **American Academy of Child and Adolescent Psychiatry** • **July 2012** • **N/A** • **https://www.aacap.org/App_Themes/AACAP/docs/practice_parameters/Atypical_Antipsychotic_Medications_Web.pdf**
Quote the guideline or recommendation verbatim	Recommendation 12:

about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	• "Careful attention should be given to the increased risk of developing diabetes with the use of AAA, and blood glucose and other parameters should be assessed at baseline and monitored at regular intervals."
Grade assigned to the **evidence** associated with the recommendation with definition of the grade	Grade of Recommendation: Minimal Standard/Clinical Standard: • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time
Provide all other grades and definitions from the evidence grading system	Grade of Recommendation: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time) Options • Acceptable but not requires; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated. AACAP Strength of Empirical Evidence: AACAP rates the strength of the empirical evidence in descending order as follows: • (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 non-randomly 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 the definition of the grade	Grade of Recommendation: Minimal Standard/Clinical Standard • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time
Provide all other grades and definitions from the recommendation grading system	Grade of Recommendation: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time) Options • Acceptable but not required; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated.
What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?	This measure addresses metabolic monitoring as one facet of safe and judicious use of antipsychotics in children and adolescents. Given the documented metabolic risks of antipsychotic medications, monitoring of metabolic indices is important to ensure appropriate management of side effect risk, especially in youth. Numerous guidelines address the need for metabolic monitoring among youth on antipsychotic medications. This measure is based on this American Academy of Child and Adolescent Psychiatry (AACAP) guideline, as well as other AACAP guidelines and by other organizations, such as the Canadian Alliance for Monitoring Effectiveness and Safety of Antipsychotics in Children (CAMESA). These organizations recommend metabolic testing for youth prescribed antipsychotics, with consensus that baseline and ongoing metabolic monitoring are standards of care for this population.
What is the time period covered by the body of evidence?	1990-2010.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Quantity: • When developing their guidelines, AACAP limited its evidence review to clinical trials, meta-analysis, practice guidelines, randomized controlled trials (RCTs), systematic literature reviews, and case reports and series. AACAP selected a total of 147 publications for careful examination based on their weight

	in the hierarchy of evidence attending to the quality of individual studies, relevance to clinical practice and the strength of the entire body of evidence. However, AACAP did not provide a breakdown of specific numbers of each publication type. Given the number of studies selected we did not feel comfortable re-conducting the evidence review and delineating all the publication types for each guideline. Instead we have identified where there are certain publication types available to support each guideline.
	• This recommendation is based on previous studies on various populations, including adults, focused on the association between diabetes/abnormal glucose regulation and the use of antipsychotics.
	• Expert opinion from four medical professional societies
	  ○ American Diabetes Association, American Psychiatric Association, American Association of Clinical Endocrinologists, North American Association for the Study of Obesity. Consensus development conference on antipsychotic drugs and obesity and diabetes. *Diabetes Care.* 2004;27:596-601.
	• Literature reviews, including case reports, case series, observational analytic epidemiologic studies, uncontrolled observations, large retrospective database analyses, and controlled experimental studies, such as randomized clinical trials
	  ○ Casey DE, Haupt DW, Newcomer JW, et al. Antipsychotic-induced weight gain and metabolic abnormalities: implications for increased mortality in patients with schizophrenia. *J Clin Psychiatry.* 2004;65[suppl 7]:4-18.
	  ○ Newcomer JW. Second-generation (atypical) antipsychotics and metabolic effects: a comprehensive literature review. *CNS Drugs.* 2005;19:1-93.
	• Case series, including five cases
	  ○ Bloch Y, Vardi O, Mendlovic S, Levkovitz Y, Gothelf D, Ratzoni G. Hyperglycemia from olanzapine treatment in adolescents. *J Child Adolesc Psychopharmacol.* 2003;13:97-102.
	• Observational study
	  ○ Hedenmalm K, Hagg S, Stahl M, Mortimer O, Spigset O. Glucose intolerance with atypical antipsychotics. *Drug Saf.* 2002;25:1107-1116.
	• Randomized clinical trial; double-blind, randomized, placebo-controlled study

	○ Henderson DC, Copeland PM, Daley TB, et al. A double-blind placebo-controlled trial of sibutramine for olanzapine associated weight gain. *Am J Psychiatry*. 2005;162:954-962. Quality: • The evidence review used by AACAP prioritized study designs less subject to bias and studies that represent the best scientific evidence. The evidence review included a large number of studies with large numbers of patients from various populations. Overall, the quality of the evidence regarding metabolic monitoring for children and adolescents on antipsychotics is high. The evidence provides a strong link between antipsychotic use and adverse metabolic side effects in youth and to negative long-term health outcomes throughout the lifespan.
Estimates of benefit and consistency across studies	AACAP did not cite any studies that directly evaluated the benefit of metabolic monitoring for children and adolescents on antipsychotics. However, the evidence demonstrates the adverse side effects, including diabetes, weight gain, and hyperlipidemia, as well as the concerns regarding the safety of long-term antipsychotics use in youth. Thus, AACAP estimates there is a greater benefit to be gained through increased vigilance and regular metabolic monitoring.
What harms were identified?	AACAP did not cite any studies that directly evaluated the harm of metabolic monitoring for children and adolescents on antipsychotics. AACAP noted that some patients and their parents may face negative social consequences due to frequent medical appointments, including greater time constraints on school and work responsibilities. However, given the adverse side effects and concerns regarding the safety of long-term antipsychotics use in youth, regular metabolic monitoring is still vital in the follow-up of these patients. Thus, in the absence of evidence of other harms, AACAP estimated there is less harm through increased vigilance and regular metabolic monitoring.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	To our knowledge, there have been no new studies that contradict the current body of evidence.

Table 3. American Academy of Child and Adolescent Psychiatry – Atypical Antipsychotic Medications 2011 Guideline on Lipid Profiles.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Practice Parameter for the Use of Atypical Antipsychotic Medications in Children and Adolescents** • **American Academy of Child and Adolescent Psychiatry** • **July 2012** • **N/A** • **https://www.aacap.org/App_Themes/AACAP/docs/practice_parameters/Atypical_Antipsychotic_Medications_Web.pdf**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Recommendation 13: • "In those patients with significant weight changes and/or a family history indicating high risk, lipid profiles should be obtained at baseline and monitored at regular intervals."
Grade assigned to the **evidence** associated with the recommendation with definition of the grade	Grade of Recommendation: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time)
Provide all other grades and definitions from the evidence grading system	Grade of Recommendation: Minimal Standard/Clinical Standard: • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time Options • Acceptable but not requires; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated. AACAP Strength of Empirical Evidence: AACAP rates the strength of the empirical evidence in descending order as follows:

	• (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 non-randomly 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 the definition of the grade	Grade of Recommendation: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time)
Provide all other grades and definitions from the recommendation grading system	Grade of Recommendation: Minimal Standard/Clinical Standard • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time Options • Acceptable but not required; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated.
What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?	This measure addresses metabolic monitoring as one facet of safe and judicious use of antipsychotics in children and adolescents. Given the documented metabolic risks of antipsychotic medications, monitoring of metabolic indices is important to ensure appropriate management of side effect risk, especially in youth. Numerous guidelines address the need for metabolic monitoring among youth on antipsychotic medications. This measure is based on this American Academy of Child and Adolescent Psychiatry (AACAP) guideline, as well as other AACAP guidelines and by other organizations, such as the Canadian Alliance for Monitoring Effectiveness and Safety of Antipsychotics in Children (CAMESA). These organizations recommend metabolic testing for youth

	prescribed antipsychotics, with consensus that baseline and ongoing metabolic monitoring are standards of care for this population.
What is the time period covered by the body of evidence?	1990-2010.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Quantity: • When developing their guidelines, AACAP limited its evidence review to clinical trials, meta-analysis, practice guidelines, randomized controlled trials (RCTs), systematic literature reviews, and case reports and series. AACAP selected a total of 147 publications for careful examination based on their weight in the hierarchy of evidence attending to the quality of individual studies, relevance to clinical practice and the strength of the entire body of evidence. However, AACAP did not provide a breakdown of specific numbers of each publication type. Given the number of studies selected we did not feel comfortable re-conducting the evidence review and delineating all the publication types for each guideline. Instead we have identified where there are certain publication types available to support each guideline. • This recommendation is based on a review of seven national, cross-sectional studies conducted between 1973 and 1994 that focused on the association between elevated lipid levels and the development of cardiovascular disease throughout the lifespan. • Freedman DS, Dietz WH, Srinivasan SR, Berenson GS., The relation of overweight to cardiovascular risk factors among children and adolescents: the Bogalusa Heart Study. *Pediatrics.* 1999;103:1175-1182. Quality: • The evidence review used by AACAP prioritized study designs less subject to bias and studies that represent the best scientific evidence. The evidence review included a large number of studies with large numbers of patients from various populations. Overall, the quality of the evidence regarding metabolic monitoring for children and adolescents on antipsychotics is high. The evidence provides a strong link between antipsychotic use and adverse metabolic side effects in youth and to negative long-term health outcomes throughout the lifespan.

Estimates of benefit and consistency across studies	AACAP did not cite any studies that directly evaluated the benefit of metabolic monitoring for children and adolescents on antipsychotics. However, the evidence demonstrates the adverse side effects, including diabetes, weight gain, and hyperlipidemia, as well as the concerns regarding the safety of long-term antipsychotics use in youth. Thus, AACAP estimates there is a greater benefit to be gained through increased vigilance and regular metabolic monitoring.
What harms were identified?	AACAP did not cite any studies that directly evaluated the harm of metabolic monitoring for children and adolescents on antipsychotics. AACAP noted that some patients and their parents may face negative social consequences due to frequent medical appointments, including greater time constraints on school and work responsibilities. However, given the adverse side effects and concerns regarding the safety of long-term antipsychotics use in youth, regular metabolic monitoring is still vital in the follow-up of these patients. Thus, in the absence of evidence of other harms, AACAP estimated there is less harm through increased vigilance and regular metabolic monitoring.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	To our knowledge, there have been no new studies that contradict the current body of evidence.

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.

N/A

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

N/A

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

N/A

1b. Performance Gap

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

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

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

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

This measure addresses metabolic monitoring as one facet of safe and judicious use of antipsychotics in children and adolescents. Although antipsychotic medications offer the potential for effective treatment of psychiatric disorders in children, they can also increase a child's risk for developing serious metabolic health complications associated with poor cardiometabolic outcomes in adulthood. Despite the risk of such adverse side effects and clinical guideline recommendation, evidence suggests that children and adolescents do not receive appropriate laboratory monitoring. Thus, this measure encourages metabolic monitoring of children who are on antipsychotic medications.

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

The following data demonstrate the variation in the rate of children and adolescents who had two or more antipsychotic prescriptions and had metabolic testing.

Metabolic Monitoring for Children and Adolescents on Antipsychotics

N = Number of Health Plans

YEAR = Measurement Year

Commercial

YEAR|N|MEAN|STDEV|MIN|10th|25th|50th|75th|90th|MAX|Interquartile Range

2016|279|34%|10%|8%|22%|27%|33%|39%|47%|80%|12%

2017|272|35%|11%|13%|22%|27%|34%|41%|48%|77%|14%

2018|273|35%|10%|12%|23%|28%|34%|40%|47%|69%|12%

Medicaid

YEAR|N|MEAN|ST DEV|MIN|10th|25th|50th|75th|90th|MAX|Interquartile Range

2016|164|33%|11%|11%|22%|25%|32%|39%|48%|70%|14%

2017|166|35%|13%|12%|22%|26%|32%|41%|51%|83%|15%

2018|169|35%|12%|6%|23%|27%|33%|41%|49%|71%|14%

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), which serves as a proxy for socioeconomic status. While not specified in the measure, this measure can also be stratified by health plans using demographic variables, such as race/ethnicity, gender or other variables, if the plan has sufficient data to support the stratifications. 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

There is little research on potential disparities in metabolic monitoring for children and adolescents prescribed antipsychotics. One study found that race/ethnicity was not associated with glucose or lipid screening rates (Morrato et al., 2010). As part of the HEDIS measure's field testing, we assessed differences in metabolic screening and monitoring in Medicaid children by race/ethnicity. Our results indicate that Hispanic children had better (i.e. higher) rates of baseline metabolic screening (10.3 percent) compared to white non-Hispanic children (5.7 percent) and black non-Hispanic children (6.1 percent). We also found that Hispanic children also had better (i.e. higher) rates of ongoing metabolic monitoring (24.8 percent) compared to white non-Hispanic children (19.1 percent) and black non-Hispanic children (19.4 percent).

Among youth receiving antipsychotics on Medicaid, there is a marked disparity in metabolic monitoring by foster care status. A 2011 study found that 28 percent of Medicaid foster children received metabolic monitoring during the year, while only 18 percent of Medicaid non-foster children received metabolic monitoring during the year (Crystal et al., 2016).

Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2011.

Crystal S, Mackie T, Fenton MC, et al. (2016). Rapid Growth of Antipsychotic Prescriptions for Children Who Are Publicly Insured Has Ceased, but Concerns Remain. Health Affairs. 2016;35(6):974-82.

Morrato E, Nicol G, Maahs D, Druss B, Hartung D, Valuck R et al. (2010). Metabolic screening in children receiving antipsychotic drug treatment. Arch Pediatr Adolesc Med, 164, 344-351.

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

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

Safety, Safety : Medication, Screening

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

Children, Populations at Risk

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

N/A

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** 2800_APM_Value_Sets_Fall_2019.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

There have been minor changes to the measure. NCQA combined the 1-5 and 6-11 year age stratifications. Separate evaluation of children 1-5 years of age is limited as utilization of antipsychotics among this age group is low. Additionally, NCQA added two rates to evaluate glucose testing and cholesterol testing independently. The total rate reflecting both glucose and cholesterol testing remains unchanged. Expert and stakeholder feedback indicate that glucose testing among children on antipsychotics is more common than cholesterol testing. Evaluation of each metabolic test type separately provides additional information to support safe and coordinated management of children on antipsychotics.

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.*

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

Children and adolescents 1-17 years of age on antipsychotics who received blood glucose and cholesterol testing during the measurement year.

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

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

Three numerators are reported using administrative data:

1. Children and adolescents 1-17 years of age on antipsychotics who received blood glucose testing during the measurement year.
2. Children and adolescents 1-17 years of age on antipsychotics who received cholesterol testing during the measurement year.
3. Children and adolescents on antipsychotics who received blood glucose and cholesterol testing during the measurement year.

Blood Glucose Testing: one test for blood glucose (Glucose Lab Test Value Set; Glucose Test Result or Finding Value Set) or HbA1c (HbA1c Lab Test Value Set; HbA1c Test Result or Finding Value Set) during the measurement year.

Cholesterol Testing: one test for LDL-C (LDL-C Lab Test Value Set; LDL-C Test Result or Finding Value Set) or cholesterol (Cholesterol Lab Test Value Set; Cholesterol Test Result or Finding Value Set) during the measurement year.

Blood Glucose and Cholesterol Testing: both of the following during the measurement year on the same or different dates of service.

- At least one test for blood glucose (Glucose Lab Test Value Set, Glucose Test Result or Finding Value Set) or HbA1c (HbA1c Lab Test Value Set, HbA1c Test Result or Finding Value Set).

- At least one test for LDL-C (LDL-C Lab Test Value Set; LDL-C Test Result or Finding Value Set) or cholesterol (Cholesterol Lab Test Value Set; Cholesterol Test Result or Finding Value Set).

See attachment for all value sets referenced above.

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

Children and adolescents 1-17 years of age who had ongoing use of antipsychotic medications (at least two prescriptions).

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

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

Children and adolescents age 1-17 years as of December 31 of the measurement year who had at least two antipsychotic medication dispensing events (Table APM-A) of the same or different medications, on different dates of service during the measurement year, with no more than one gap in enrollment of up to 45 days during the measurement year.

TABLE APM-A: ANTIPSYCHOTIC MEDICATIONS

DESCRIPTION / PRESCRIPTION

Miscellaneous antipsychotic agents / Aripiprazole; Asenapine; Brexpiprazole; Cariprazine; Clozapine; Haloperidol; Iloperidone; Loxapine; Lurisadone; Molindone; Olanzapine; Paliperidone; Pimozide; Quetiapine; Quetiapine fumarate, Risperidone, Ziprasidone

Phenothiazine antipsychotics / Chlorpromazine; Fluphenazine; Perphenazine; Thioridazine; Trifluoperazine

Thioxanthenes / Thiothixene

Long-acting injections / Aripiprazole; Fluphenazine decanoate; Haloperidol decanoate; Olanzapine; Paliperidone palmitate; Risperidone

Psychotherapeutics combinations / Fluoxetine-olanzapine; Perphenazine-amitriptyline

Phenothiazine antipsychotics / Prochlorperazine

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

Patients in hospice.

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

Exclude 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 patients may be identified using various methods, which may include but are not limited to enrollment data, medical record or claims/encounter data (Hospice Encounter Value Set or Hospice Intervention Value Set).

See corresponding Excel file for value sets referenced above.

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.)*

Report two age stratifications and a total rate:

• Children and adolescents 1-11 years of age as of December 31 of the measurement year.

• Children and adolescents 12-17 years of age as of December 31 of the measurement year.

• Total (the sum of the age stratifications).

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

STEP 1: Determine the eligible population. To do so, identify patients who meet all the specified criteria.

- AGES: Children and adolescents 1-17 years of age as of December 31 of the measurement year.
- EVENT/DIAGNOSIS: Identify patients who had at least two antipsychotic medication dispensing events of the same or different medications, on different dates of service during the measurement year. SEE S.7 for the list of antipsychotic medications.

STEP 2: Determine the numerator by identifying the number of patients in the eligible population who received blood glucose testing, cholesterol testing, or blood glucose testing and cholesterol testing.

STEP 3: Calculate the rate by dividing the numerator by the denominator.

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

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

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

Emergency Department and Services, 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

APM_Testing_Form_-2800-_updated_11.14.19.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*)**:** 2800
Measure Title: Metabolic Monitoring for Children and Adolescents on Antipsychotics
Date of Submission: 8/1/2019

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*).

2019 Submission: This measure was tested using administrative claims data from Medicaid and commercial plans nationwide that reported data for the annual Healthcare Effectiveness Data and Information Set (HEDIS®[1]).

2015 Submission: This measure was tested using administrative claims data from the following sources.

- State analyses
  - Medicaid Analytic eXtract (MAX)
- Health plan analyses
  - Medicaid health plans from one state
  - Sample of commercial health plans nationwide

For more information about MAX, refer to http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Data-and-Systems/MAX/MAX-General-Information.html

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

2019 Submission: Testing of performance measure score with beta binomial reliability and testing of construct validity with the Pearson Correlation were performed using HEDIS plan level data, measurement year 2018.

2015 Submission: MAX data 2008, Medicaid health plan data for 17 plans 2010, and commercial health plan data for 73 plans 2012.

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

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

[1] HEDIS is a registered trademark of the National Committee for Quality Assurance

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*)

2019 Submission: *Metabolic Monitoring for Children and Adolescents on Antipsychotics* is a HEDIS health-plan level measure that assesses whether youth on antipsychotics receive blood glucose and cholesterol testing during the year. We tested the measure at the health plan level, which is appropriate for the accountable entity for which the measure is specified. We calculated the measure score reliability and construct validity from HEDIS data that included 273 commercial plans and 169 Medicaid plans. The sample included all commercial and Medicaid health plans submitting data to NCQA for this HEDIS measure. The plans were geographically diverse and varied in size.

2015 Submission: During measure development: As part of the Pediatric Quality Measures Program (PQMP), National Collaborative for Innovation in Quality Measurement (NCINQ) had access to the Medicaid Analytic eXtract (MAX) for conducting state analyses. In addition, NCINQ was able to test this measure in Medicaid health plan data from one large mid-Atlantic state. In order to assess the measure's use for HEDIS, we conducted an additional analysis in commercial data from a large administrative database. Our samples were as follows.

- State analyses
    - 2008 claims data from the MAX for 11 states
- Health plan analyses
    - 2010 claims data from 17 Medicaid health plans from one mid-Atlantic state
    - 2012 claims data from 73 commercial health plans nationwide

The administrative data sources included claims for all of the data elements needed to capture this measure, including claims for health care system encounters, laboratory codes, and pharmacy codes.

For our MAX analysis, the 11 states were chosen on the basis of Mathematica Policy Research reports that suggested that they provided adequate encounter/managed care data (Byrd & Dodd, 2012; Byrd & Dodd, 2013).

Byrd VLH, Dodd AH. Assessing the usability of encounter data for enrollees in comprehensive managed care across MAX 2007-2009. December 2012.

Byrd VLH, Dodd AH. Assessing the Usability of MAX 2008 Encounter Data for Comprehensive Managed Care. Medicare & Medicaid Research Review. 2013;3(1).

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*)

2019 Submission: Below is a description of the data submitted for 2018, including the mean denominator size per plan. The denominator for this measure includes children and adolescents age 1 to

17 years with at least two dispensing events for antipsychotics during the measurement year. Data are summarized at the health plan level and stratified by plan type (i.e. commercial, Medicaid).

Table 1. Mean denominator size per plan for Metabolic Monitoring for Children and Adolescents on Antipsychotics, 2018.

Product Line	Number of Plans	Mean Denominator Size/Plan
Commercial	273	245
Medicaid	169	1128

2015 Submission: We tested a set of several measures related to antipsychotic use in the three datasets described above. Our analyses included enrollees who met continuous enrollment and measure-specific criteria. Our commercial health plan analyses included enrollees age 0-17 years during the measurement year. All other analyses included enrollees ages 0 to 20 during the measurement year. The age ranges varied slightly as our draft concepts were refined and in order to make the measures relevant to states (children/adolescents typically defined as age up to 21 years) and health plans (children/adolescents typically defined as age up to 18 years). We excluded enrollees who were dually eligible for Medicaid and Medicare. In the MAX data, a total of 148,910 children and adolescents met the denominator criteria and were included in the sample for this measure. Across the 17 Medicaid plans, the total number of children and adolescents who met denominator criteria was 14,174, and across 52 commercial plans that had sufficient denominators (>30), the total was 15,227.

Below are descriptions of the patient samples in terms of denominator sizes across the entities measured. They include the mean denominator, minimum denominator, maximum denominator, and the 25th, 50th (or median), and 75th percentiles.

Denominator Size Distribution Across 11 States (MAX) (2008)

Mean	**13,537**
Minimum	1,784
25th	6,272
Median	12,372
75th	18,684
Maximum	28,997

Denominator Size Distribution Across 17 State Medicaid Health Plans from One State (2010)

Mean	**834**
Minimum	125
25th	306
Median	748

75th	1,082
Maximum	2,437

Denominator Size Distribution Across 52* Commercial Health Plans Nationwide (2012)

Mean	**293**
Minimum	33
25th	103
Median	206
75th	369
Maximum	1,870

* Of the 73 commercial plans included in the testing of this measure, 52 had sufficient denominators (>30)

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.

2019 Submission:

Reliability:

Reliability of the measure score was tested using a beta-binomial calculation. This analysis included the entire 2018 data reported by health plans for HEDIS (described above).

Validity:

Validity of the health plan measure was demonstrated through a systematic assessment of face validity and construct validity using the entire HEDIS data sample (described above).

2015 Submission: Reliability of the measure score was tested using a beta-binomial calculation and this analysis included the entire data samples described in the sections above (MAX state data, Medicaid health plan, commercial health plan).

Validity was demonstrated through a systematic assessment of face validity. Per NQF instructions we have described the composition of the technical expert panels which assessed face validity in the data sample questions above. In addition, validity was demonstrated through two types of analyses: correlations among measures using Pearson Correlation Coefficients (using commercial health plan data sample) and rankings of health plans and states on measures (using MAX state data sample and Medicaid health plan data sample). This analysis is described further in section 2b1.3.

For identifying statistically significant & meaningful differences in performance, all three data samples were used (MAX state data, Medicaid health plan, commercial health plan).

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.

2019 and 2015 Submission: We did not analyze social risk factors. This measure of health plan performance is specified to be reported separately by Medicaid and commercial plan types, which serves as a proxy for income and other socioeconomic factors.

2a2. RELIABILITY TESTING

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

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

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

2019 and 2015 Submission:

We utilized the Beta-binomial method (Adams, 2009) to assess how well one can confidently distinguish between performance across accountable entities. Conceptually, the Beta-binomial measures the proportion of total variation attributable to a health plan, which represents the "signal" and also estimates the proportion of variation attributable to measurement error for each plan, which represents "noise." The reliability of the measure is represented as the ratio of signal to noise. A score of 0 indicates none of the variation (signal) is attributable to the plan. A score of 1.0 indicates all of the variation (signal) is attributable to the plan. A score of 0.7 or higher indicates adequate reliability to distinguish performance between two plans.

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

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

2019 Submission: Table 2 shows the reliability for the overall measure as shown by the Beta-binomial model as well as the distribution of individual plan reliability.

Table 2. Overall Beta-binomial statistic and distribution of plan reliability for commercial and Medicaid product lines, 2018.

Product Line	Overall Reliability	Min	Percentiles					Max
			10th	25th	50th	75th	90th	

Commercial	0.875	0.439	0.540	0.644	0.795	0.892	0.948	0.985
Medicaid	0.985	0.648	0.864	0.936	0.975	0.990	0.994	0.998

2015 Submission: This measure achieved a reliability score above 0.7 for both state- and plan-level reliability.

Data Source	Overall Reliability	Minimum Reliability
MAX States	.99	.99
Medicaid Health Plan	.98	.89
Commercial Health Plan	.83	.35

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?*)
2019 Submission: The values for the overall beta-binomial statistic across all product lines for the health plan level measure are all greater than 0.7, indicating the measure has very good reliability. The distribution of health plan level-reliability on this measure shows that approximately half of commercial health plans and nearly all Medicaid health plans exceeded the 0.7 threshold for adequate reliability. Good reliability is demonstrated since most variance is due to signal and not to noise.

2015 Submission: The testing results suggest that this measure has adequate reliability for states and health plans, with very high reliability for Medicaid health plans and states in particular.

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)*

2019 Submission:

Method of Testing Construct Validity

We tested for construct validity of the Metabolic Monitoring for Children and Adolescents on Antipsychotics measure by exploring whether it was correlated with other similar measures of quality hypothesized which are listed below.

- Adolescent Well-Care Visits: The percentage of enrolled patients 12 to 21 years of age who had at least one comprehensive well-care visit with a primary care physician (PCP) or
- Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life: The percentage of patients 3 to 6 years of age who had one or more well-child visits with a primary care physician (PCP) during the measurement year.

These measures were chosen for construct validity testing because these measures capture the annual well-care screening and monitoring that should be provided to youth, similar to the routine metabolic monitoring that should be performed for youth on antipsychotics. We hypothesized that organizations that perform well on the HEDIS Well Care measures should perform well on the Metabolic Monitoring measure.

To test these correlations, we continued usage of the Pearson correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable. Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less that this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

2015 Submission:

Method of Testing Construct Validity

We tested for construct validity by exploring whether this measure was correlated with other related measures, including the Metabolic Screening for Children and Adolescents on Antipsychotics measure, the Follow-up Visit for Children and Adolescents on Antipsychotics measure and the Use of First-line Psychosocial Care for Children and Adolescents on Antipsychotics measure. The Metabolic Screening measure assesses the percentage of youth who undergo metabolic testing prior to or immediately after the start of a new antipsychotic prescription. The Follow-up Visit measure assesses the percentage of youth who have one or more visits with a prescriber within 30 days after the start of a new antipsychotic prescription. The Psychosocial Care measure assesses the percentage of youth who have psychosocial care provided before or soon after the start of a new antipsychotic prescription. A higher rate indicates better performance for all three measures.

Method of Assessing Face Validity

The health-plan level of this measure was assessed for use in the HEDIS Health Plan Measure Set. As part of this process, NCQA assessed the face validity of the measure using its HEDIS process. NCQA staff shared the measure concepts, supporting evidence and field test results with its standing Behavioral Health Measurement Advisory Panel, Technical Measurement Advisory Panel and additional panels. We posted the measures for Public Comment, a 30-day period of review that allowed interested parties to

offer feedback about the measure. NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations.

NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle. This measure has undergone the following steps associated with that cycle.

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

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

Step 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA and the CPM about new measures or about changes to existing measures. NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM and NCQA's Board of Directors will be included in the next HEDIS year and reported as first-year measures.

Step 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's State of Health Care Quality, Quality Compass or in accreditation scoring. The first-year distinction guarantees that a measure can be effectively collected, reported and audited before it is used for public accountability or accreditation. This is not testing—the measure was already tested as part of its development—rather, it ensures that there are no unforeseen problems when the measure is implemented in the real world. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

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

Step 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reviewed for reevaluation at least every three years. NCQA

staff continually monitors the performance of publicly reported measures. Statistical analysis, audit results review, and user comments through NCQA's Policy Clarification Support portal contribute to measure refinement during re-evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

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

** Note: All HEDIS value sets are updated annually with the most current codes available. The information below details the process we used to convert value sets that used ICD-9 codes to ICD-10 codes in 2015. **

ICD-10 CONVERSION:

In preparation for the national implementation of ICD-10 in 2015, NCQA conducted a systematic mapping of all value sets maintained by the organization to ensure the new values used for reporting maintained the reliability, validity and intent of the original specification.

Steps in ICD-9 to ICD-10 Conversion Process

1. NCQA staff identified ICD-10 codes to be considered based on ICD-9 codes currently in measure. Used GEM to identify ICD-10 codes that map to ICD-9 codes. Reviewed GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
2. NCQA staff identified additional codes (not identified by GEM mapping step) that should be considered. Using ICD-10 tabular list and ICD-10 Index, searched by diagnosis or procedure name for appropriate codes.
3. NCQA HEDIS Expert Coding Panel reviewed NCQA staff recommendations and provided feedback.
4. As needed, NCQA Measurement Advisory Panels performed clinical review. Due to increased specificity in ICD-10, new codes and definitions required review to confirm the diagnosis or procedure was intended to be included in the scope of the measure. Not all ICD-10 recommendations were reviewed by NCQA MAP; MAP review items were identified during staff conversion or by HEDIS Expert Coding Panel.
5. Posted ICD-10 code recommendations for public review and comment.
6. Reconciled public comments. Obtained additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
7. NCQA staff finalized ICD-10 code recommendations.

Tools Used to Identify/Map to ICD-10

All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).

GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

Expert Participation

The NCQA HEDIS Expert Coding Panel reviewed and provided feedback on staff recommendations. Names and credentials of the experts who served on these panels are listed under Additional Information, Ad. 1. Workgroup/Expert Panel Involved in Measure Development.

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

2019 Submission:

Construct Validity

The results from construct validity testing of the health plan level measure are presented by product line in Tables 3a, 3b, and 3c below.

Table 3a. Results of Pearson Correlation Coefficient for commercial health plans, 2018.

Measure	**Pearson Correlation Coefficients**	
	Adolescent Well-Care Visits	Well-Child Visit
Metabolic Monitoring	0.41	0.37

**includes data submitted by 273 Commercial plans to HEDIS for these measures for measurement year 2018*

Note: all correlations are significant at p<.0001

Table 3b. Results of Pearson Correlation Coefficient for Medicaid health plans, 2018.

Measure	**Pearson Correlation Coefficients**	
	Adolescent Well-Care Visits	Well-Child Visit
Metabolic Monitoring	0.31	0.39

**includes data submitted by 169 Commercial plans to HEDIS for these measures for measurement year 2018*

Note: all correlations are significant at p<.0001

2015 Submission:

Construct Validity

When determining correlations among measures, we focused on health plans, as there were not enough entities to measure correlations with the state data.

Among national commercial plans, there was a very slight positive correlation between the *First-line Psychosocial Care* and *Metabolic Monitoring* measures (r=0.12, p=.70) and high positive correlation between the *Metabolic Screening* and *Metabolic Monitoring* measures (r=0.82, p<0.0001).

Measure	**Pearson Correlation Coefficients**		
	Psychosocial Care	Metabolic Screening	Metabolic Monitoring
Psychosocial Care	1	0.18	0.12
Metabolic Screening		1	0.82
Metabolic Monitoring			1

Among Medicaid plans in one state, there was a slight positive correlation between the *Follow-up Visit* and *Metabolic Monitoring* measures (r=0.14, p=.58) and high positive correlation between the *Metabolic Screening* and *Metabolic Monitoring* measures (r=0.72, p<0.001).

Ranking

Among MAX states and one state's Medicaid plans, we found good consistency in the states and plans, respectively, with the best and worst performance.

MAX State Performance Rankings: General Population

State	Metabolic Monitoring	Metabolic Screening	Follow-Up Visit
1	14.2	**2.6**	**60.2**
2	19.4	4.5	**68.4**
3	20.6	5.5	75.0
4	**6.5**	3.8	71.2
5	**4.8**	**0.4**	74.9
6	18.7	4.8	76.4
7	20.0	6.3	69.0
8	14.8	5.3	N/A
9	**29.1**	**10.7**	N/A
10	19.6	8.3	**81.3**
11	**36.2**	**14.0**	**78.8**
Mean	**18.5**	**6.0**	**72.8**

Medicaid Health Plan Performance Rankings for One State

Plan	Metabolic Monitoring	Metabolic Screening	Follow-Up Visit
3	**2.3**	**0.2**	**71.0**
9	30.8	**4.9**	81.8
6	34.0	12.3	83.5
17	**39.7**	14.8	**86.7**
2	38.8	**15.4**	80.5
8	35.0	12.6	81.1
4	28.4	9.3	78.7
5	33.8	10.6	80.0
1	36.0	12.8	82.1
11	29.1	6.1	74.4
16	31.2	10.6	78.8

15	30.4	10.8	80.9
12	34.7	13.3	77.2
13	32.5	**17.8**	**70.4**
7	**20.3**	5.1	85.3
14	27.9	7.1	**98.7**
10	**40.0**	10.6	78.9
Mean	**30.9**	**10.3**	**80.6**

Face Validity

Step 1: This measure was developed to address the need for metabolic monitoring of children and adolescents who are on antipsychotics. NCQA and five expert panels worked together in 2013 and 2014 to identify the most appropriate method for assessing metabolic monitoring among this patient population. Across the multiple expert panels that reviewed this measure, all panels concluded this measure was specified to assess metabolic monitoring.

Step 2: The measure was written and field-tested in 2013 and 2014. After reviewing field test results, the CPM recommended to send the measure to public comment with a majority vote in January 2014.

Step 3: The measure was released for Public Comment in 2014 prior to publication in HEDIS. Of 67 comments received, nearly all (94 percent) supported it as-is or with suggested modifications. The CPM recommended moving this measure to first year data collection by a majority vote in May 2014.

Step 4: The measure was introduced in HEDIS 2015. Organizations voluntarily reported this measure in the first year (2014) and the results were analyzed for public reporting in the following year (2015). The measure was approved in September 2015 by the CPM for public reporting in HEDIS 2016 for commercial and Medicaid plans.

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?*)

2019 Submission:

Construct Validity

Across all product lines, the correlations are moderate and statistically significant. These results confirm the hypothesis that plan performance on these Well-Visit measures are correlated to performance on the Metabolic Monitoring for Children and Adolescents on Antipsychotics measure. Plans that have higher rates on one measure will have higher rates on the other. Coefficients with absolute value of less than .3 are generally considered indicative of weak associations. Absolute values of .3 to .59 are considered moderate associations, absolute values of .6 to .69 indicate a strong positive relationship, and absolute values of .7 or higher indicate a very strong positive relationship. These correlation results suggest that at the plan level the measure has sufficient validity.

2015 Submission:

Construct Validity

Correlations

Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone. The results indicate that plans that perform well on initial metabolic screening for those youth newly on antipsychotics also perform well on ongoing metabolic monitoring for those who continue on antipsychotics.

Ranking

The results show that plans and states can be approximately ranked based on profiles of performance across multiple measures. The consistent performance across measures suggest the measures are assessing a dimension of quality.

Face Validity

The results indicate the expert panels showed good agreement that the measure as specified will accurately differentiate quality across health plans. Our interpretation of these results is that this measure has sufficient face validity.

2b2. EXCLUSIONS ANALYSIS

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

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

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

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

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

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

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

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with** 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)*

2019 and 2015 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 .05, then the two plans' performance is significantly different from each other.

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*)

2019 Submission:

Table 4. Variation in Performance for Commercial and Medicaid health plans, 2018.

Plan Type	N	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value

Commercial	273	34.7%	9.6%	23.3%	28.2%	33.8%	40.4%	46.8%	12.2%	<0.0001
Medicaid	169	35.3%	12.2%	23.1%	27.4%	33.3%	40.9%	49.1%	13.5%	<0.0001

N = Number of plans reporting

IQR = Interquartile range

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

2015 Submission:

Variation in Performance Rates across MAX States (2008 data)

Mean Rate	10th	25th	50th	75th	90th	IQR
18.5	6.5	14.2	19.4	20.6	29.1	6.4

IQR: Interquartile range

Variation in Performance Rates across Medicaid Plans from one State (2010 data)

Mean Rate	10th	25th	50th	75th	90th	IQR
30.9	24.9	28.8	32.5	35.5	39.2	6.7

IQR: Interquartile range

Variation in Performance Rates across Commercial Plans Nationwide (2012 data)

Mean Rate	10th	25th	50th	75th	90th	IQR
7.7	2.8	4.5	7.2	10.9	13.2	6.4

IQR: Interquartile range

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?*)
2019 Submission: The results above indicate there is meaningful difference in performance. For both commercial and Medicaid plans, the difference between the 25th and 75th percentile (better performance) is statistically significant.

2015 Submission: The results indicate that there is 6.7% gap in performance between Medicaid plans at the 25th and 75th percentiles, a 6.4% gap in performance among commercial plans and a 6.4% gap in performance among states at the 25th and 75th percentiles. This means that states at the 25th percentile have on average 866 less children and adolescents getting recommended metabolic monitoring than states at the 75th percentile. For Medicaid plans, those at the 25th percentile have on average 56 less children and adolescents getting recommended metabolic monitoring than plans at the 75th percentile. For commercial plans, those at the 25th percentile have on average 19 less children and adolescents getting recommended metabolic monitoring than plans at the 75th percentile.

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

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

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

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

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

2019 Submission:

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

The HEDIS Compliance Audit addresses the following functions:

- Information practices and control procedures
- Sampling methods and procedures

- Data integrity
- Compliance with HEDIS specifications
- Analytic file production
- Reporting and documentation

2015 Submission:

States and plans collect this measure using all administrative data sources, for all intents and purposes, there are no missing data in administrative data. We have done no assessment to look for the distribution of missing data. For plans reporting on this measure for HEDIS, NCQA's audit process checks that plans' measure calculations are not biased due to missing data.

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*)
2019 and 2015 Submission: HEDIS addresses missing data in a structured way through its audit process. HEDIS measures apply to enrolled members in a health plan, and NCQA-certified auditors use standard audit methodologies to assess whether data sources are missing data. If a data source is found to be missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used. Once measures are added to HEDIS, NCQA conducts a first-year analysis to assess the measure's feasibility once widely implemented in the field. This analysis includes an assessment of how many plans report valid rates vs. rates that are materially biased (or have other issues, such as small denominators). These considerations are weighed in the deliberation process before measures are approved for public reporting.

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*)
2019 and 2015 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.

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).

N/A

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

Attachment:

3c. Data Collection Strategy

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

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

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

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

The HEDIS Compliance Audit addresses the following functions:

1) Information practices and control procedures

2) Sampling methods and procedures

3) Data integrity

4) Compliance with HEDIS specifications

5) Analytic file production

6) Reporting and documentation

In addition to the HEDIS Audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measures. This system is vital to the regular re-evaluation of the NCQA measures.

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 scientific soundness 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)*.

N/A

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 Rating https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ Annual State of Health Care Quality: https://www.ncqa.org/hedis/measures/metabolic-monitoring-for-children-and-adolescents-on-antipsychotics/ Health Plan Rating https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ Annual State of Health Care Quality: https://www.ncqa.org/hedis/measures/metabolic-monitoring-for-children-and-adolescents-on-antipsychotics/ Regulatory and Accreditation Programs Health Plan Accreditation https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ Quality Improvement (external benchmarking to organizations) Quality Compass https://www.ncqa.org/programs/data-and-information-technology/data-purchase-and-licensing/quality-compass/ Annual State of Health Care Quality: https://www.ncqa.org/hedis/measures/metabolic-monitoring-for-children-and-adolescents-on-antipsychotics/

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

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 2018, the report included results from calendar year 2017 for health plans covering a record 136 million people, or 43 percent of the U.S. population.
HEALTH PLAN RATING/REPORT CARDS: This measure is used to calculate health plan ratings which are reported in Consumer Reports and on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. In 2019, a total of 538 Medicare Advantage health plans, 544 commercial health plans and 268 Medicaid health plans across 50 states were included in the ratings.
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.

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 Health Care Quality Congress, NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly. NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System, as described in Section 3c.1.

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

Describe how feedback was obtained.

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

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

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.

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

This measure has been deemed a priority measure by NCQA and other entities, as illustrated by its use in programs such as the Annual State of Healthcare Quality and the Health Plan Rating.

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.

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.

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.

Performance on this measure has improved slightly for Medicaid and commercial health plans (see section 1b.2 for summary of data for commercial and Medicaid health plans). The largest improvement in performance scores is seen in the Medicaid product line. In 2019, a total of 273 commercial health plans and 169 Medicaid health plans across 50 states reported calendar year 2018 data on this measure. These data are nationally representative.

4b2. Unintended Consequences

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

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

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

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.

Yes

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

1932 : Diabetes Screening for People With Schizophrenia or Bipolar Disorder Who Are Using Antipsychotic Medications (SSD)

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

N/A

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?

No

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

The Metabolic Monitoring for Children and Adolescents on Antipsychotics measure assesses metabolic monitoring during the measurement year among children and adolescents who are prescribed antipsychotics. This measure is related to measure #1932 but addresses a different target population and measure focus. Measure #1932 assesses whether adults with schizophrenia or bipolar disorder who were prescribed antipsychotics are screened for diabetes. Similar to the Metabolic Monitoring for Children and Adolescents on Antipsychotics measure, this measure is specified for the health plan level and uses administrative claims as the data source. The measures have different target populations but a similar measure focus. Measure #1932 focuses on adults 18 to 64 years of age who have schizophrenia or bipolar disorder and who are prescribed antipsychotics. The Metabolic Monitoring for Children and Adolescents on Antipsychotics measure includes all children and adolescents up to 17 years of age who are prescribed antipsychotics and does not focus on any specific conditions. Measure #1932 is focused on diabetes screening by receipt of a glucose test. While the Metabolic Monitoring for Children and Adolescents on Antipsychotics measure also includes assessing whether a glucose test was received, it additionally assesses whether a cholesterol test was received since the focus is not just diabetes screening. The two measures are aligned in the way glucose testing is identified and measured.

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, 202-955-3500-

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

Co.4 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-3500-

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.

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, CenCalHealth

Harmon Jordan, ScD, Westat

Virginia Raney, LCSW, Center for Medicaid and CHIP Services

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

Laurie Spoll, Aetna

NATIONAL COLLABORATIVE FOR INNOVATION IN QUALITY MEASUREMENT (NCINQ) MEASUREMENT ADVISORY PANEL

Samantha Broderick, Cardinal Catalysts

Cathy Caldwell, MPH, Alabama Department of Public Health

Jim Chase, MHA, Network for Regional Healthcare Improvement

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

Kim Elliott, PhD, Health Services Advisory Group

Lynda Gargan, PhD, National Federation of Families for Children's Mental Health

David Harmon, Md, FAAP, Superior Health Plan

Jennifer Havens, MD, NYU Langone Health

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

Sheree Neese-Todd, MS, Rutgers University, Center for Health Services Research

Rob Penfold, PhD, Kaiser Permanente Washington Health Research Institute

Cheryl J. Roberts, JD, Virginia Department of Medical Assistance Services

Jeff Schiff, MD, MBA, Minnesota Department of Human Services

Julie Sonier, MPA, MN Community Measurement

Craig Thiele, MD, Health Management Associates

Jeb Weisman, PhD, Icahn School of Medicine at Mount Sinai

Judy Zerzan, MD, MPH, WA Health Care Authority

NCINQ CONSUMER PANEL

Joan Alker, MPhil, Georgetown Center for Children and Families

Roni Christopher, MEd, OTR/L, PCMH-CCE, The Greater Cincinnati Health Collaborative

Daniel Coury, MD, Nationwide Children's Hospital

Eileen Forlenza, Colorado Medical Home Initiative, Children and Youth with Special Health Care Needs Unit

Michaelle Gady, JD, Families USA

Janis Guerney, JD, Family Voices

Jocelyn Guyer, MPA, Georgetown Center for Children and Families

Catherine Hess, MSW, National Academy for State Health Policy

Carolyn Muller, RN, Montgomery County Health Department

Cindy Pellegrini, March of Dimes

Judith Shaw, EdD, MPH, RN, VCHIP

Stuart Spielman, JD, LLM, Autism Speaks

Michelle Sternthal, PhD, March of Dimes

NCINQ FOSTER CARE PANEL

Kamala Allen, MHS, Center for Health Care Strategies

Mary Applegate, MD, Ohio Department of Job and Family Services

Samantha Jo Broderick, Foster Care Alumni of America

Mary Greiner, MD, Cincinnati Children's Hospital Medical Center

David Harmon, MD, FAAP, Superior HealthPlan

Patricia Hunt, Magellan Health Services

Audrey LaFrenier, MSW, Parsons Child and Family Center

Bryan Samuels, MPP, Chapin Hall

Phil Scribano, DO, MSCE , The Children's Hospital of Philadelphia

Lesley Siegel, MD, State of Connecticut Department of Children and Families

Chauncey Strong, MSW, LGSW, Fairfax County Department of Family Services/Foster Care and Adoption

Janet (Jessie) Sullivan, MD, Hudson Health Plan

Nora Wells, MS, National Center for Family/Professional Partnerships

NCINQ MENTAL HEALTH PANEL

Francisca Azocar, PhD, Optum Health Behavioral Solutions

Frank Ghinassi, PhD, Western Psychiatric Institute and Clinic of UPMC Presbyterian Shadyside

Jennifer Havens, MD, NYU Langone Health

Danielle Laraque, MD, FAAP, Maimonides Infants and Children's Hospital of Brooklyn

NCINQ STATE PANEL

Mary Applegate, MD, Ohio Department of Job and Family Services

Susan Castellano, Minnesota Department of Human Services

Sharon Carte, MHS, State of West Virginia Children's Health Insurance Program

Catherine Hess, MSW, National Academy for State Health Policy

Michael Hogan, PhD, New York State office of Mental Health

Barbara Lantz, MN, RN, State of Washington Department of Social and Health Services, Medicaid Purchasing Administration

Judy Mohr Peterson, PhD, Oregon Health Authority

Tracy Plouck, MPA, Ohio Department of Mental Health

Gina Robinson, Colorado Department of Health Care Policy and Financing

Janet Stover, Illinois Association of Rehabilitation Facilities

Eric Trupin, PhD, University of Washington

NCINQ MEASURE DEVELOPMENT PARTNERS

Shahla Amin, MS, Rutgers University

Scott Bilder, PhD, Center for Health Services Research, Rutgers University

Stephen Crystal, PhD, Institute for Health, Health Care Policy and Aging Research, Rutgers University

Molly Finnerty, PhD, NY State Office of Mental Health

Emily Leckman-Westin, PhD, NY State Office of Mental Health

Sheree Neese-Todd, MA, Institute for Health, Health Care Policy and Aging Research, Rutgers University

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2014

Ad.3 Month and Year of most recent revision: 07, 2019

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? 06, 2020

Calculated measure results, based on unadjusted HEDIS specifications, may not be termed "Health Plan HEDIS rates" until they are audited and designated reportable by an NCQA-Certified Auditor. Such unaudited results should be referred to as "Unaudited Health Plan HEDIS Rates." Accordingly, "Heath Plan HEDIS rate" refers to and assumes a result from an unadjusted HEDIS specification that has been audited by an NCQA-Certified HEDIS Auditor.

Limited proprietary coding is contained in the measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. NCQA disclaims all liability for use or accuracy of any coding contained in the specifications.

Content reproduced with permission from HEDIS, Volume 2: Technical Specifications for Health Plans. To purchase copies of this publication, including the full measures and specifications, contact NCQA Customer Support at 888-275-7585 or visit www.ncqa.org/publications.

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

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

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures, without modification, are encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Modifications to, and/or commercial use of, a measure requires the prior written consent of NCQA and is subject to a license at the discretion of NCQA. As used herein, "commercial use" refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

Brief Measure Information

NQF #: 2800

Corresponding Measures:

De.2. **Measure Title:** Metabolic Monitoring for Children and Adolescents on Antipsychotics

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

De.3. **Brief Description of Measure:** The percentage of children and adolescents 1-17 years of age who had two or more antipsychotic prescriptions and had metabolic testing.

1b.1. **Developer Rationale:** This measure addresses metabolic monitoring as one facet of safe and judicious use of antipsychotics in children and adolescents. Although antipsychotic medications offer the potential for effective treatment of psychiatric disorders in children, they can also increase a child's risk for developing serious metabolic health complications associated with poor cardiometabolic outcomes in adulthood. Despite the risk of such adverse side effects and clinical guideline recommendation, evidence suggests that children and adolescents do not receive appropriate laboratory monitoring. Thus, this measure encourages metabolic monitoring of children who are on antipsychotic medications.

S.4. **Numerator Statement:** Children and adolescents 1-17 years of age on antipsychotics who received blood glucose and cholesterol testing during the measurement year.

S.6. **Denominator Statement:** Children and adolescents 1-17 years of age who had ongoing use of antipsychotic medications (at least two prescriptions). S.8. **Denominator Exclusions:** Patients in hospice.
De.1. **Measure Type:** Process S.17. **Data Source:** Claims S.20. **Level of Analysis:** Health Plan
IF Endorsement Maintenance – Original Endorsement Date: May 04, 2016 **Most Recent Endorsement Date:** May 04, 2016
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

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

APM_Evidence_Form_-2800-.docx

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

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

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2800

Measure Title: Metabolic Monitoring for Children and Adolescents on Antipsychotics

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

Date of Submission: 8/1/2019

Instructions

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

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

1a. Evidence to Support the Measure Focus

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

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

Notes

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

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

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

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

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

Outcome

☐ Outcome: Click here to name the health outcome

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

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

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

☒ Process: Annual blood glucose and cholesterol testing for children and adolescents on antipsychotics

☐ 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.

This measure assesses metabolic monitoring (i.e., the receipt of glucose and cholesterol tests) among children and adolescents that have ongoing antipsychotic use. Given the documented metabolic risks of antipsychotic medications, monitoring of metabolic indices is important to ensure appropriate management of side effect risk, especially in youth. The path envisioned is as follows.

Child or adolescent has ongoing use of antipsychotic medication >>> Metabolic monitoring by a health care provider >>> Identification of metabolic issues/side effects >>> Health care provider addresses metabolic issue >>> Patient receives intervention for metabolic issues present >>> Metabolic issues reduced or eliminated >>> Improvement in metabolic functioning for patient (desired outcome).

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

N/A

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

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

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 Child and Adolescent Psychiatry – Atypical Antipsychotic Medications 2011 Guideline on Fasting Profiles.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Practice Parameter for the Use of Atypical Antipsychotic Medications in Children and Adolescents** • **American Academy of Child and Adolescent Psychiatry** • **July 2012** • **N/A** • **https://www.aacap.org/App_Themes/AACAP/docs/practice_parameters/Atypical_Antipsychotic_Medications_Web.pdf**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Table: Fasting plasma glucose – Baseline, 12 wks, annually; Fasting lipid profile – Baseline, 12 wks (Recommendation 10, and Table 2): • "The acute and long-term safety of these medications in children and adolescents has not been fully evaluated and therefore careful and frequent monitoring of side effects should be performed…*Ideally, monitoring of BMI, blood pressure, fasting glucose and fasting lipid profiles should follow, whenever feasible, the recommendations found in the consensus statement put forth by the American Diabetes Association and American Psychiatric Association*."

Grade assigned to the **evidence** associated with the recommendation with definition of the grade	Grade of Recommendation: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time)
Provide all other grades and definitions from the evidence grading system	Grade of Recommendation: Minimal Standard/Clinical Standard: • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time Options • Acceptable but not requires; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated. AACAP Strength of Empirical Evidence: AACAP rates the strength of the empirical evidence in descending order as follows: • (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 non-randomly 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 the definition of the grade	Grade of Recommendation: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time)

Provide all other grades and definitions from the recommendation grading system	Grade of Recommendation: Minimal Standard/Clinical Standard • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time Options • Acceptable but not required; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated.
What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?	This measure addresses metabolic monitoring as one facet of safe and judicious use of antipsychotics in children and adolescents. Given the documented metabolic risks of antipsychotic medications, monitoring of metabolic indices is important to ensure appropriate management of side effect risk, especially in youth. Numerous guidelines address the need for metabolic monitoring among youth on antipsychotic medications. This measure is based on this American Academy of Child and Adolescent Psychiatry (AACAP) guideline, as well as other AACAP guidelines and by other organizations, such as the Canadian Alliance for Monitoring Effectiveness and Safety of Antipsychotics in Children (CAMESA). These organizations recommend metabolic testing for youth prescribed antipsychotics, with consensus that baseline and ongoing metabolic monitoring are standards of care for this population.
What is the time period covered by the body of evidence?	1990-2010.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Quantity: • When developing their guidelines, AACAP limited its evidence review to clinical trials, meta-analysis, practice guidelines, randomized controlled trials (RCTs), systematic literature reviews, and case reports and series. AACAP selected a total of 147 publications for careful examination based on their weight in the hierarchy of evidence attending to the quality of individual studies, relevance to clinical practice and the strength of the entire body of evidence. However, AACAP did not provide a breakdown of specific numbers of each publication type. Given the number of studies selected we did not feel comfortable re-conducting the evidence review and delineating all the publication types for each guideline. Instead

	we have identified where there are certain publication types available to support each guideline. • This recommendation is based on expert opinion established during a consensus development conference for four medical professional societies. The four societies found that an increasing number of methodologically rigorous studies have assessed the effectiveness of antipsychotics for children and adolescents in specific clinical situations. However, the long-term safety profile of each antipsychotic used by youth has yet to be effectively evaluated and characterized. In the absence of such evidence, AACAP recommends increased vigilance. • American Diabetes Association, American Psychiatric Association, American Association of Clinical Endocrinologists, North American Association for the Study of Obesity. Consensus development conference on antipsychotic drugs and obesity and diabetes. *Diabetes Care.* 2004;27:596-601. Quality: • The evidence review used by AACAP prioritized study designs less subject to bias and studies that represent the best scientific evidence. The evidence review included a large number of studies with large numbers of patients from various populations. Overall, the quality of the evidence regarding metabolic monitoring for children and adolescents on antipsychotics is high. The evidence provides a strong link between antipsychotic use and adverse metabolic side effects in youth and to negative long-term health outcomes throughout the lifespan.
Estimates of benefit and consistency across studies	AACAP did not cite any studies that directly evaluated the benefit of metabolic monitoring for children and adolescents on antipsychotics. However, the evidence demonstrates the adverse side effects, including diabetes, weight gain, and hyperlipidemia, as well as the concerns regarding the safety of long-term antipsychotics use in youth. Thus, AACAP estimates there is a greater benefit to be gained through increased vigilance and regular metabolic monitoring.
What harms were identified?	AACAP did not cite any studies that directly evaluated the harm of metabolic monitoring for children and adolescents on antipsychotics. AACAP noted that some patients and their parents may face negative social consequences due to frequent medical appointments, including greater time constraints on school and work responsibilities. However, given the adverse side effects and concerns regarding the safety of long-term antipsychotics use in youth, regular metabolic monitoring is

	still vital in the follow-up of these patients. Thus, in the absence of evidence of other harms, AACAP estimated there is less harm through increased vigilance and regular metabolic monitoring.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	To our knowledge, there have been no new studies that contradict the current body of evidence.

Table 2. American Academy of Child and Adolescent Psychiatry – Atypical Antipsychotic Medications 2011 Guideline on Blood Glucose Monitoring.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Practice Parameter for the Use of Atypical Antipsychotic Medications in Children and Adolescents** • **American Academy of Child and Adolescent Psychiatry** • **July 2012** • **N/A** • **https://www.aacap.org/App_Themes/AACAP/docs/practice_parameters/Atypical_Antipsychotic_Medications_Web.pdf**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Recommendation 12: • "Careful attention should be given to the increased risk of developing diabetes with the use of AAA, and blood glucose and other parameters should be assessed at baseline and monitored at regular intervals."
Grade assigned to the **evidence** associated with the recommendation with definition of the grade	Grade of Recommendation: Minimal Standard/Clinical Standard: • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time
Provide all other grades and definitions from the evidence grading system	Grade of Recommendation: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time) Options

	• Acceptable but not requires; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated. <u>AACAP Strength of Empirical Evidence</u>: AACAP rates the strength of the empirical evidence in descending order as follows: • (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 non-randomly 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 the definition of the grade	<u>Grade of Recommendation</u>: Minimal Standard/Clinical Standard • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time
Provide all other grades and definitions from the recommendation grading system	<u>Grade of Recommendation</u>: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time) Options • Acceptable but not required; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated.

What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?	This measure addresses metabolic monitoring as one facet of safe and judicious use of antipsychotics in children and adolescents. Given the documented metabolic risks of antipsychotic medications, monitoring of metabolic indices is important to ensure appropriate management of side effect risk, especially in youth. Numerous guidelines address the need for metabolic monitoring among youth on antipsychotic medications. This measure is based on this American Academy of Child and Adolescent Psychiatry (AACAP) guideline, as well as other AACAP guidelines and by other organizations, such as the Canadian Alliance for Monitoring Effectiveness and Safety of Antipsychotics in Children (CAMESA). These organizations recommend metabolic testing for youth prescribed antipsychotics, with consensus that baseline and ongoing metabolic monitoring are standards of care for this population.
What is the time period covered by the body of evidence?	1990-2010.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Quantity: • When developing their guidelines, AACAP limited its evidence review to clinical trials, meta-analysis, practice guidelines, randomized controlled trials (RCTs), systematic literature reviews, and case reports and series. AACAP selected a total of 147 publications for careful examination based on their weight in the hierarchy of evidence attending to the quality of individual studies, relevance to clinical practice and the strength of the entire body of evidence. However, AACAP did not provide a breakdown of specific numbers of each publication type. Given the number of studies selected we did not feel comfortable re-conducting the evidence review and delineating all the publication types for each guideline. Instead we have identified where there are certain publication types available to support each guideline. • This recommendation is based on previous studies on various populations, including adults, focused on the association between diabetes/abnormal glucose regulation and the use of antipsychotics. • Expert opinion from four medical professional societies ○ American Diabetes Association, American Psychiatric Association, American Association of Clinical Endocrinologists, North American Association for the Study of Obesity. Consensus development conference on antipsychotic drugs and obesity and diabetes. *Diabetes Care.* 2004;27:596-601. • Literature reviews, including case reports, case series, observational analytic epidemiologic studies, uncontrolled

	observations, large retrospective database analyses, and controlled experimental studies, such as randomized clinical trials ○ Casey DE, Haupt DW, Newcomer JW, et al. Antipsychotic-induced weight gain and metabolic abnormalities: implications for increased mortality in patients with schizophrenia. *J Clin Psychiatry.* 2004;65[suppl 7]:4-18. ○ Newcomer JW. Second-generation (atypical) antipsychotics and metabolic effects: a comprehensive literature review. *CNS Drugs.* 2005;19:1-93. • Case series, including five cases ○ Bloch Y, Vardi O, Mendlovic S, Levkovitz Y, Gothelf D, Ratzoni G. Hyperglycemia from olanzapine treatment in adolescents. *J Child Adolesc Psychopharmacol.* 2003;13:97-102. • Observational study ○ Hedenmalm K, Hagg S, Stahl M, Mortimer O, Spigset O. Glucose intolerance with atypical antipsychotics. *Drug Saf.* 2002;25:1107-1116. • Randomized clinical trial; double-blind, randomized, placebo-controlled study ○ Henderson DC, Copeland PM, Daley TB, et al. A double-blind placebo-controlled trial of sibutramine for olanzapine associated weight gain. *Am J Psychiatry.* 2005;162:954-962. Quality: • The evidence review used by AACAP prioritized study designs less subject to bias and studies that represent the best scientific evidence. The evidence review included a large number of studies with large numbers of patients from various populations. Overall, the quality of the evidence regarding metabolic monitoring for children and adolescents on antipsychotics is high. The evidence provides a strong link between antipsychotic use and adverse metabolic side effects in youth and to negative long-term health outcomes throughout the lifespan.
Estimates of benefit and consistency across studies	AACAP did not cite any studies that directly evaluated the benefit of metabolic monitoring for children and adolescents on antipsychotics. However, the evidence demonstrates the adverse side effects, including diabetes, weight gain, and hyperlipidemia, as well as the concerns regarding the safety of long-term antipsychotics use in youth. Thus,

	AACAP estimates there is a greater benefit to be gained through increased vigilance and regular metabolic monitoring.
What harms were identified?	AACAP did not cite any studies that directly evaluated the harm of metabolic monitoring for children and adolescents on antipsychotics. AACAP noted that some patients and their parents may face negative social consequences due to frequent medical appointments, including greater time constraints on school and work responsibilities. However, given the adverse side effects and concerns regarding the safety of long-term antipsychotics use in youth, regular metabolic monitoring is still vital in the follow-up of these patients. Thus, in the absence of evidence of other harms, AACAP estimated there is less harm through increased vigilance and regular metabolic monitoring.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	To our knowledge, there have been no new studies that contradict the current body of evidence.

Table 3. American Academy of Child and Adolescent Psychiatry – Atypical Antipsychotic Medications 2011 Guideline on Lipid Profiles.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Practice Parameter for the Use of Atypical Antipsychotic Medications in Children and Adolescents** • **American Academy of Child and Adolescent Psychiatry** • **July 2012** • **N/A** • **https://www.aacap.org/App_Themes/AACAP/docs/practice_parameters/Atypical_Antipsychotic_Medications_Web.pdf**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Recommendation 13: • "In those patients with significant weight changes and/or a family history indicating high risk, lipid profiles should be obtained at baseline and monitored at regular intervals."
Grade assigned to the **evidence** associated with the recommendation with definition of the grade	Grade of Recommendation: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time)

Provide all other grades and definitions from the evidence grading system	Grade of Recommendation: Minimal Standard/Clinical Standard: • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time Options • Acceptable but not requires; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated. AACAP Strength of Empirical Evidence: AACAP rates the strength of the empirical evidence in descending order as follows: • (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 non-randomly 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 the definition of the grade	Grade of Recommendation: Clinical Guideline • Strong empirical evidence (nonrandomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expect to apply in most cases (75 percent of the time)
Provide all other grades and definitions from the recommendation grading system	Grade of Recommendation: Minimal Standard/Clinical Standard • Rigorous/substantial empirical evidence (meta-analyses, systematic reviews, RCTs) and/or overwhelming clinical consensus; expected to apply more than 95 percent of the time

	Options • Acceptable but not required; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports). Not Endorsed • Ineffective or contraindicated.
What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?	This measure addresses metabolic monitoring as one facet of safe and judicious use of antipsychotics in children and adolescents. Given the documented metabolic risks of antipsychotic medications, monitoring of metabolic indices is important to ensure appropriate management of side effect risk, especially in youth. Numerous guidelines address the need for metabolic monitoring among youth on antipsychotic medications. This measure is based on this American Academy of Child and Adolescent Psychiatry (AACAP) guideline, as well as other AACAP guidelines and by other organizations, such as the Canadian Alliance for Monitoring Effectiveness and Safety of Antipsychotics in Children (CAMESA). These organizations recommend metabolic testing for youth prescribed antipsychotics, with consensus that baseline and ongoing metabolic monitoring are standards of care for this population.
What is the time period covered by the body of evidence?	1990-2010.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Quantity: • When developing their guidelines, AACAP limited its evidence review to clinical trials, meta-analysis, practice guidelines, randomized controlled trials (RCTs), systematic literature reviews, and case reports and series. AACAP selected a total of 147 publications for careful examination based on their weight in the hierarchy of evidence attending to the quality of individual studies, relevance to clinical practice and the strength of the entire body of evidence. However, AACAP did not provide a breakdown of specific numbers of each publication type. Given the number of studies selected we did not feel comfortable re-conducting the evidence review and delineating all the publication types for each guideline. Instead we have identified where there are certain publication types available to support each guideline. • This recommendation is based on a review of seven national, cross-sectional studies conducted between 1973 and 1994 that focused on the association between elevated lipid levels and

	the development of cardiovascular disease throughout the lifespan. • Freedman DS, Dietz WH, Srinivasan SR, Berenson GS., The relation of overweight to cardiovascular risk factors among children and adolescents: the Bogalusa Heart Study. *Pediatrics.* 1999;103:1175-1182. Quality: • The evidence review used by AACAP prioritized study designs less subject to bias and studies that represent the best scientific evidence. The evidence review included a large number of studies with large numbers of patients from various populations. Overall, the quality of the evidence regarding metabolic monitoring for children and adolescents on antipsychotics is high. The evidence provides a strong link between antipsychotic use and adverse metabolic side effects in youth and to negative long-term health outcomes throughout the lifespan.
Estimates of benefit and consistency across studies	AACAP did not cite any studies that directly evaluated the benefit of metabolic monitoring for children and adolescents on antipsychotics. However, the evidence demonstrates the adverse side effects, including diabetes, weight gain, and hyperlipidemia, as well as the concerns regarding the safety of long-term antipsychotics use in youth. Thus, AACAP estimates there is a greater benefit to be gained through increased vigilance and regular metabolic monitoring.
What harms were identified?	AACAP did not cite any studies that directly evaluated the harm of metabolic monitoring for children and adolescents on antipsychotics. AACAP noted that some patients and their parents may face negative social consequences due to frequent medical appointments, including greater time constraints on school and work responsibilities. However, given the adverse side effects and concerns regarding the safety of long-term antipsychotics use in youth, regular metabolic monitoring is still vital in the follow-up of these patients. Thus, in the absence of evidence of other harms, AACAP estimated there is less harm through increased vigilance and regular metabolic monitoring.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	To our knowledge, there have been no new studies that contradict the current body of evidence.

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.

N/A

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

N/A

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

N/A

1b. Performance Gap

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

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.

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

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

Measure Number (*if previously endorsed*)**:** 2800
Measure Title: Metabolic Monitoring for Children and Adolescents on Antipsychotics
Date of Submission: 8/1/2019

Type of Measure:

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

Instructions

- **Measures must be tested for all the data sources and levels of analyses that are specified. *If there is more than one set of data specifications or more than one level of analysis, contact NQF staff* about how to present all the testing information in one form.**
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures, section 2b3 also must be completed.**
- **If specified for multiple data sources/sets of specificaitons (e.g., claims and EHRs), section 2b5 also must be completed.**
- **Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.**
- **If you are unable to check a box, please highlight or shade the box for your response.**
- **Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). *Contact NQF staff if more pages are needed*.**
- **Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.**
- **For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.**

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

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

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

- **an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration**

OR

- **rationale/data support no risk adjustment/ stratification.**

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

OR

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

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

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

Notes

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

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

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

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

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

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g.,

$5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

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

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

2019 Submission: This measure was tested using administrative claims data from Medicaid and commercial plans nationwide that reported data for the annual Healthcare Effectiveness Data and Information Set (HEDIS®[2]).

2015 Submission: This measure was tested using administrative claims data from the following sources.

- State analyses
  - Medicaid Analytic eXtract (MAX)
- Health plan analyses
  - Medicaid health plans from one state
  - Sample of commercial health plans nationwide

For more information about MAX, refer to http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Data-and-Systems/MAX/MAX-General-Information.html

[2] HEDIS is a registered trademark of the National Committee for Quality Assurance

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

2019 Submission: Testing of performance measure score with beta binomial reliability and testing of construct validity with the Pearson Correlation were performed using HEDIS plan level data, measurement year 2018.

2015 Submission: MAX data 2008, Medicaid health plan data for 17 plans 2010, and commercial health plan data for 73 plans 2012.

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

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

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

2019 Submission: *Metabolic Monitoring for Children and Adolescents on Antipsychotics* is a HEDIS health-plan level measure that assesses whether youth on antipsychotics receive blood glucose and cholesterol testing during the year. We tested the measure at the health plan level, which is appropriate for the accountable entity for which the measure is specified. We calculated the measure score reliability and construct validity from HEDIS data that included 273 commercial plans and 169 Medicaid plans. The sample included all commercial and Medicaid health plans submitting data to NCQA for this HEDIS measure. The plans were geographically diverse and varied in size.

2015 Submission: During measure development: As part of the Pediatric Quality Measures Program (PQMP), National Collaborative for Innovation in Quality Measurement (NCINQ) had access to the Medicaid Analytic eXtract (MAX) for conducting state analyses. In addition, NCINQ was able to test this measure in Medicaid health plan data from one large mid-Atlantic state. In order to assess the measure's use for HEDIS, we conducted an additional analysis in commercial data from a large administrative database. Our samples were as follows.

- State analyses
    - 2008 claims data from the MAX for 11 states

- Health plan analyses
    - 2010 claims data from 17 Medicaid health plans from one mid-Atlantic state
    - 2012 claims data from 73 commercial health plans nationwide

The administrative data sources included claims for all of the data elements needed to capture this measure, including claims for health care system encounters, laboratory codes, and pharmacy codes.

For our MAX analysis, the 11 states were chosen on the basis of Mathematica Policy Research reports that suggested that they provided adequate encounter/managed care data (Byrd & Dodd, 2012; Byrd & Dodd, 2013).

Byrd VLH, Dodd AH. Assessing the usability of encounter data for enrollees in comprehensive managed care across MAX 2007-2009. December 2012.

Byrd VLH, Dodd AH. Assessing the Usability of MAX 2008 Encounter Data for Comprehensive Managed Care. Medicare & Medicaid Research Review. 2013;3(1).

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*)

2019 Submission: Below is a description of the data submitted for 2018, including the mean denominator size per plan. The denominator for this measure includes children and adolescents age 1 to 17 years with at least two dispensing events for antipsychotics during the measurement year. Data are summarized at the health plan level and stratified by plan type (i.e. commercial, Medicaid).

Table 1. Mean denominator size per plan for Metabolic Monitoring for Children and Adolescents on Antipsychotics, 2018.

Product Line	Number of Plans	Mean Denominator Size/Plan
Commercial	273	245
Medicaid	169	1128

2015 Submission: We tested a set of several measures related to antipsychotic use in the three datasets described above. Our analyses included enrollees who met continuous enrollment and measure-specific criteria. Our commercial health plan analyses included enrollees age 0-17 years during the measurement year. All other analyses included enrollees ages 0 to 20 during the measurement year. The age ranges varied slightly as our draft concepts were refined and in order to make the measures relevant to states (children/adolescents typically defined as age up to 21 years) and health plans (children/adolescents typically defined as age up to 18 years). We excluded enrollees who were dually eligible for Medicaid and Medicare. In the MAX data, a total of 148,910 children and adolescents met the denominator criteria and were included in the sample for this measure. Across the 17 Medicaid plans, the total

number of children and adolescents who met denominator criteria was 14,174, and across 52 commercial plans that had sufficient denominators (>30), the total was 15,227.

Below are descriptions of the patient samples in terms of denominator sizes across the entities measured. They include the mean denominator, minimum denominator, maximum denominator, and the 25th, 50th (or median), and 75th percentiles.

Denominator Size Distribution Across 11 States (MAX) (2008)

Mean	**13,537**
Minimum	1,784
25th	6,272
Median	12,372
75th	18,684
Maximum	28,997

Denominator Size Distribution Across 17 State Medicaid Health Plans from One State (2010)

Mean	**834**
Minimum	125
25th	306
Median	748
75th	1,082
Maximum	2,437

Denominator Size Distribution Across 52* Commercial Health Plans Nationwide (2012)

Mean	**293**
Minimum	33
25th	103
Median	206
75th	369
Maximum	1,870

* Of the 73 commercial plans included in the testing of this measure, 52 had sufficient denominators (>30)

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.

2019 Submission:

Reliability:

Reliability of the measure score was tested using a beta-binomial calculation. This analysis included the entire 2018 data reported by health plans for HEDIS (described above).

Validity:

Validity of the health plan measure was demonstrated through a systematic assessment of face validity and construct validity using the entire HEDIS data sample (described above).

2015 Submission: Reliability of the measure score was tested using a beta-binomial calculation and this analysis included the entire data samples described in the sections above (MAX state data, Medicaid health plan, commercial health plan).

Validity was demonstrated through a systematic assessment of face validity. Per NQF instructions we have described the composition of the technical expert panels which assessed face validity in the data sample questions above. In addition, validity was demonstrated through two types of analyses: correlations among measures using Pearson Correlation Coefficients (using commercial health plan data sample) and rankings of health plans and states on measures (using MAX state data sample and Medicaid health plan data sample). This analysis is described further in section 2b1.3.

For identifying statistically significant & meaningful differences in performance, all three data samples were used (MAX state data, Medicaid health plan, commercial health plan).

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.

2019 and 2015 Submission: We did not analyze social risk factors. This measure of health plan performance is specified to be reported separately by Medicaid and commercial plan types, which serves as a proxy for income and other socioeconomic factors.

2a2. RELIABILITY TESTING

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

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

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

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

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

2019 and 2015 Submission:

We utilized the Beta-binomial method (Adams, 2009) to assess how well one can confidently distinguish between performance across accountable entities. Conceptually, the Beta-binomial measures the proportion of total variation attributable to a health plan, which represents the "signal" and also estimates the proportion of variation attributable to measurement error for each plan, which represents "noise." The reliability of the measure is represented as the ratio of signal to noise. A score of 0 indicates none of the variation (signal) is attributable to the plan. A score of 1.0 indicates all of the variation (signal) is attributable to the plan. A score of 0.7 or higher indicates adequate reliability to distinguish performance between two plans.

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

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

2019 Submission: Table 2 shows the reliability for the overall measure as shown by the Beta-binomial model as well as the distribution of individual plan reliability.

Table 2. Overall Beta-binomial statistic and distribution of plan reliability for commercial and Medicaid product lines, 2018.

Product Line	Overall Reliability	Min	Percentiles					Max
			10th	25th	50th	75th	90th	
Commercial	0.875	0.439	0.540	0.644	0.795	0.892	0.948	0.985
Medicaid	0.985	0.648	0.864	0.936	0.975	0.990	0.994	0.998

2015 Submission: This measure achieved a reliability score above 0.7 for both state- and plan-level reliability.

Data Source	Overall Reliability	Minimum Reliability
MAX States	.99	.99
Medicaid Health Plan	.98	.89
Commercial Health Plan	.83	.35

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?*)

2019 Submission: The values for the overall beta-binomial statistic across all product lines for the health plan level measure are all greater than 0.7, indicating the measure has very good reliability. The distribution of health plan level-reliability on this measure shows that approximately half of commercial

health plans and nearly all Medicaid health plans exceeded the 0.7 threshold for adequate reliability. Good reliability is demonstrated since most variance is due to signal and not to noise.

2015 Submission: The testing results suggest that this measure has adequate reliability for states and health plans, with very high reliability for Medicaid health plans and states in particular.

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)*

2019 Submission:

Method of Testing Construct Validity

We tested for construct validity of the Metabolic Monitoring for Children and Adolescents on Antipsychotics measure by exploring whether it was correlated with other similar measures of quality hypothesized which are listed below.

- Adolescent Well-Care Visits: The percentage of enrolled patients 12 to 21 years of age who had at least one comprehensive well-care visit with a primary care physician (PCP) or
- Well-Child Visits in the Third, Fourth, Fifth and Sixth Years of Life: The percentage of patients 3 to 6 years of age who had one or more well-child visits with a primary care physician (PCP) during the measurement year.

These measures were chosen for construct validity testing because these measures capture the annual well-care screening and monitoring that should be provided to youth, similar to the routine metabolic monitoring that should be performed for youth on antipsychotics. We hypothesized that organizations that perform well on the HEDIS Well Care measures should perform well on the Metabolic Monitoring measure.

To test these correlations, we continued usage of the Pearson correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 and +1. A value of 1 indicates a perfect linear dependence in increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable. Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or

higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less that this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone.

2015 Submission:

Method of Testing Construct Validity

We tested for construct validity by exploring whether this measure was correlated with other related measures, including the Metabolic Screening for Children and Adolescents on Antipsychotics measure, the Follow-up Visit for Children and Adolescents on Antipsychotics measure and the Use of First-line Psychosocial Care for Children and Adolescents on Antipsychotics measure. The Metabolic Screening measure assesses the percentage of youth who undergo metabolic testing prior to or immediately after the start of a new antipsychotic prescription. The Follow-up Visit measure assesses the percentage of youth who have one or more visits with a prescriber within 30 days after the start of a new antipsychotic prescription. The Psychosocial Care measure assesses the percentage of youth who have psychosocial care provided before or soon after the start of a new antipsychotic prescription. A higher rate indicates better performance for all three measures.

Method of Assessing Face Validity

The health-plan level of this measure was assessed for use in the HEDIS Health Plan Measure Set. As part of this process, NCQA assessed the face validity of the measure using its HEDIS process. NCQA staff shared the measure concepts, supporting evidence and field test results with its standing Behavioral Health Measurement Advisory Panel, Technical Measurement Advisory Panel and additional panels. We posted the measures for Public Comment, a 30-day period of review that allowed interested parties to offer feedback about the measure. NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations.

NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle. This measure has undergone the following steps associated with that cycle.

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

Step 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential

measures. The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

Step 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA and the CPM about new measures or about changes to existing measures. NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM and NCQA's Board of Directors will be included in the next HEDIS year and reported as first-year measures.

Step 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's State of Health Care Quality, Quality Compass or in accreditation scoring. The first-year distinction guarantees that a measure can be effectively collected, reported and audited before it is used for public accountability or accreditation. This is not testing—the measure was already tested as part of its development—rather, it ensures that there are no unforeseen problems when the measure is implemented in the real world. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

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

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

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

** Note: All HEDIS value sets are updated annually with the most current codes available. The information below details the process we used to convert value sets that used ICD-9 codes to ICD-10 codes in 2015. **

ICD-10 CONVERSION:

In preparation for the national implementation of ICD-10 in 2015, NCQA conducted a systematic mapping of all value sets maintained by the organization to ensure the new values used for reporting maintained the reliability, validity and intent of the original specification.

Steps in ICD-9 to ICD-10 Conversion Process

8. NCQA staff identified ICD-10 codes to be considered based on ICD-9 codes currently in measure. Used GEM to identify ICD-10 codes that map to ICD-9 codes. Reviewed GEM mapping in both directions (ICD-9 to ICD-10 and ICD-10 to ICD-9) to identify potential trending issues.
9. NCQA staff identified additional codes (not identified by GEM mapping step) that should be considered. Using ICD-10 tabular list and ICD-10 Index, searched by diagnosis or procedure name for appropriate codes.
10. NCQA HEDIS Expert Coding Panel reviewed NCQA staff recommendations and provided feedback.
11. As needed, NCQA Measurement Advisory Panels performed clinical review. Due to increased specificity in ICD-10, new codes and definitions required review to confirm the diagnosis or procedure was intended to be included in the scope of the measure. Not all ICD-10 recommendations were reviewed by NCQA MAP; MAP review items were identified during staff conversion or by HEDIS Expert Coding Panel.
12. Posted ICD-10 code recommendations for public review and comment.
13. Reconciled public comments. Obtained additional feedback from HEDIS Expert Coding Panel and MAPs as needed.
14. NCQA staff finalized ICD-10 code recommendations.

Tools Used to Identify/Map to ICD-10

All tools used for mapping/code identification from CMS ICD-10 website (http://www.cms.gov/Medicare/Coding/ICD10/2012-ICD-10-CM-and-GEMs.html).

GEM, ICD-10 Guidelines, ICD-10-CM Tabular List of Diseases and Injuries, ICD-10-PCS Tabular List.

Expert Participation

The NCQA HEDIS Expert Coding Panel reviewed and provided feedback on staff recommendations. Names and credentials of the experts who served on these panels are listed under Additional Information, Ad. 1. Workgroup/Expert Panel Involved in Measure Development.

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

2019 Submission:

Construct Validity

The results from construct validity testing of the health plan level measure are presented by product line in Tables 3a, 3b, and 3c below.

Table 3a. Results of Pearson Correlation Coefficient for commercial health plans, 2018.

Measure	**Pearson Correlation Coefficients**	
	Adolescent Well-Care Visits	Well-Child Visit
Metabolic Monitoring	0.41	0.37

includes data submitted by 273 Commercial plans to HEDIS for these measures for measurement year 2018

Note: all correlations are significant at p<.0001

Table 3b. Results of Pearson Correlation Coefficient for Medicaid health plans, 2018.

Measure	Pearson Correlation Coefficients	
	Adolescent Well-Care Visits	Well-Child Visit
Metabolic Monitoring	0.31	0.39

includes data submitted by 169 Commercial plans to HEDIS for these measures for measurement year 2018

Note: all correlations are significant at p<.0001

2015 Submission:

Construct Validity

When determining correlations among measures, we focused on health plans, as there were not enough entities to measure correlations with the state data.

Among national commercial plans, there was a very slight positive correlation between the *First-line Psychosocial Care* and *Metabolic Monitoring* measures (r=0.12, p=.70) and high positive correlation between the *Metabolic Screening* and *Metabolic Monitoring* measures (r=0.82, p<0.0001).

Measure	Pearson Correlation Coefficients		
	Psychosocial Care	Metabolic Screening	Metabolic Monitoring
Psychosocial Care	1	0.18	0.12
Metabolic Screening		1	0.82
Metabolic Monitoring			1

Among Medicaid plans in one state, there was a slight positive correlation between the *Follow-up Visit* and *Metabolic Monitoring* measures (r=0.14, p=.58) and high positive correlation between the *Metabolic Screening* and *Metabolic Monitoring* measures (r=0.72, p<0.001).

Ranking

Among MAX states and one state's Medicaid plans, we found good consistency in the states and plans, respectively, with the best and worst performance.

MAX State Performance Rankings: General Population

State	Metabolic Monitoring	Metabolic Screening	Follow-Up Visit
1	14.2	**2.6**	**60.2**
2	19.4	4.5	**68.4**
3	20.6	5.5	75.0
4	**6.5**	3.8	71.2
5	**4.8**	**0.4**	74.9

6	18.7	4.8	76.4
7	20.0	6.3	69.0
8	14.8	5.3	N/A
9	**29.1**	**10.7**	N/A
10	19.6	8.3	**81.3**
11	**36.2**	**14.0**	**78.8**
Mean	**18.5**	**6.0**	**72.8**

Medicaid Health Plan Performance Rankings for One State

Plan	Metabolic Monitoring	Metabolic Screening	Follow-Up Visit
3	**2.3**	**0.2**	**71.0**
9	30.8	**4.9**	81.8
6	34.0	12.3	83.5
17	**39.7**	14.8	**86.7**
2	38.8	**15.4**	80.5
8	35.0	12.6	81.1
4	28.4	9.3	78.7
5	33.8	10.6	80.0
1	36.0	12.8	82.1
11	29.1	6.1	74.4
16	31.2	10.6	78.8
15	30.4	10.8	80.9
12	34.7	13.3	77.2
13	32.5	**17.8**	**70.4**
7	**20.3**	5.1	85.3
14	27.9	7.1	**98.7**
10	**40.0**	10.6	78.9
Mean	**30.9**	**10.3**	**80.6**

Face Validity

Step 1: This measure was developed to address the need for metabolic monitoring of children and adolescents who are on antipsychotics. NCQA and five expert panels worked together in 2013 and 2014 to identify the most appropriate method for assessing metabolic monitoring among this patient population. Across the multiple expert panels that reviewed this measure, all panels concluded this measure was specified to assess metabolic monitoring.

Step 2: The measure was written and field-tested in 2013 and 2014. After reviewing field test results, the CPM recommended to send the measure to public comment with a majority vote in January 2014.

Step 3: The measure was released for Public Comment in 2014 prior to publication in HEDIS. Of 67 comments received, nearly all (94 percent) supported it as-is or with suggested modifications. The CPM recommended moving this measure to first year data collection by a majority vote in May 2014.

Step 4: The measure was introduced in HEDIS 2015. Organizations voluntarily reported this measure in the first year (2014) and the results were analyzed for public reporting in the following year (2015). The measure was approved in September 2015 by the CPM for public reporting in HEDIS 2016 for commercial and Medicaid plans.

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?*)

2019 Submission:

Construct Validity

Across all product lines, the correlations are moderate and statistically significant. These results confirm the hypothesis that plan performance on these Well-Visit measures are correlated to performance on the Metabolic Monitoring for Children and Adolescents on Antipsychotics measure. Plans that have higher rates on one measure will have higher rates on the other. Coefficients with absolute value of less than .3 are generally considered indicative of weak associations. Absolute values of .3 to .59 are considered moderate associations, absolute values of .6 to .69 indicate a strong positive relationship, and absolute values of .7 or higher indicate a very strong positive relationship. These correlation results suggest that at the plan level the measure has sufficient validity.

2015 Submission:

Construct Validity

Correlations

Coefficients with absolute value of less than 0.3 are generally considered indicative of weak associations whereas absolute values of 0.3 or higher denote moderate to strong associations. The significance of a correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p-value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. We used a threshold of 0.05 to evaluate the test results. P-values less than this threshold imply that it is unlikely that a non-zero coefficient was observed due to chance alone. The results indicate that plans that perform well on initial metabolic screening for those youth newly on antipsychotics also perform well on ongoing metabolic monitoring for those who continue on antipsychotics.

Ranking

The results show that plans and states can be approximately ranked based on profiles of performance across multiple measures. The consistent performance across measures suggest the measures are assessing a dimension of quality.

Face Validity

The results indicate the expert panels showed good agreement that the measure as specified will accurately differentiate quality across health plans. Our interpretation of these results is that this measure has sufficient face validity.

2b2. EXCLUSIONS ANALYSIS

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

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

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

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

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

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

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

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with** 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)*

2019 and 2015 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 .05, then the two plans' performance is significantly different from each other.

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*)

2019 Submission:

Table 4. Variation in Performance for Commercial and Medicaid health plans, 2018.

Plan Type	N	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Commercial	273	34.7%	9.6%	23.3%	28.2%	33.8%	40.4%	46.8%	12.2%	<0.0001
Medicaid	169	35.3%	12.2%	23.1%	27.4%	33.3%	40.9%	49.1%	13.5%	<0.0001

N = Number of plans reporting

IQR = Interquartile range

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

2015 Submission:

Variation in Performance Rates across MAX States (2008 data)

Mean Rate	10th	25th	50th	75th	90th	IQR
18.5	6.5	14.2	19.4	20.6	29.1	6.4

IQR: Interquartile range

Variation in Performance Rates across Medicaid Plans from one State (2010 data)

Mean Rate	10th	25th	50th	75th	90th	IQR

30.9	24.9	28.8	32.5	35.5	39.2	6.7

IQR: Interquartile range

Variation in Performance Rates across Commercial Plans Nationwide (2012 data)

Mean Rate	10th	25th	50th	75th	90th	IQR
7.7	2.8	4.5	7.2	10.9	13.2	6.4

IQR: Interquartile range

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?*)
2019 Submission: The results above indicate there is meaningful difference in performance. For both commercial and Medicaid plans, the difference between the 25th and 75th percentile (better performance) is statistically significant.

2015 Submission: The results indicate that there is 6.7% gap in performance between Medicaid plans at the 25th and 75th percentiles, a 6.4% gap in performance among commercial plans and a 6.4% gap in performance among states at the 25th and 75th percentiles. This means that states at the 25th percentile have on average 866 less children and adolescents getting recommended metabolic monitoring than states at the 75th percentile. For Medicaid plans, those at the 25th percentile have on average 56 less children and adolescents getting recommended metabolic monitoring than plans at the 75th percentile. For commercial plans, those at the 25th percentile have on average 19 less children and adolescents getting recommended metabolic monitoring than plans at the 75th percentile.

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

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

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

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

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

2019 Submission:

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

The HEDIS Compliance Audit addresses the following functions:

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

2015 Submission:

States and plans collect this measure using all administrative data sources, for all intents and purposes, there are no missing data in administrative data. We have done no assessment to look for the distribution of missing data. For plans reporting on this measure for HEDIS, NCQA's audit process checks that plans' measure calculations are not biased due to missing data.

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*)

2019 and 2015 Submission: HEDIS addresses missing data in a structured way through its audit process. HEDIS measures apply to enrolled members in a health plan, and NCQA-certified auditors use standard audit methodologies to assess whether data sources are missing data. If a data source is found to be

missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used. Once measures are added to HEDIS, NCQA conducts a first-year analysis to assess the measure's feasibility once widely implemented in the field. This analysis includes an assessment of how many plans report valid rates vs. rates that are materially biased (or have other issues, such as small denominators). These considerations are weighed in the deliberation process before measures are approved for public reporting.

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*)

2019 and 2015 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.

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).

N/A

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. <u>Required for maintenance of endorsement.</u> 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.

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

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

The HEDIS Compliance Audit addresses the following functions:

1) Information practices and control procedures

2) Sampling methods and procedures

3) Data integrity

4) Compliance with HEDIS specifications

5) Analytic file production

6) Reporting and documentation

In addition to the HEDIS Audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measures. This system is vital to the regular re-evaluation of the NCQA measures.

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 scientific soundness 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)*.

N/A

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 Rating https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ Annual State of Health Care Quality: https://www.ncqa.org/hedis/measures/metabolic-monitoring-for-children-and-adolescents-on-antipsychotics/ Health Plan Rating https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ Annual State of Health Care Quality: https://www.ncqa.org/hedis/measures/metabolic-monitoring-for-children-and-adolescents-on-antipsychotics/ Regulatory and Accreditation Programs Health Plan Accreditation https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ Quality Improvement (external benchmarking to organizations) Quality Compass https://www.ncqa.org/programs/data-and-information-technology/data-purchase-and-licensing/quality-compass/ Annual State of Health Care Quality: https://www.ncqa.org/hedis/measures/metabolic-monitoring-for-children-and-adolescents-on-antipsychotics/

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

- Name of program and sponsor
- Purpose

- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

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 2018, the report included results from calendar year 2017 for health plans covering a record 136 million people, or 43 percent of the U.S. population.
HEALTH PLAN RATING/REPORT CARDS: This measure is used to calculate health plan ratings which are reported in Consumer Reports and on the NCQA website. These ratings are based on performance on HEDIS measures among other factors. In 2019, a total of 538 Medicare Advantage health plans, 544 commercial health plans and 268 Medicaid health plans across 50 states were included in the ratings.
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.

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 Health Care Quality Congress, NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly. NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System, as described in Section 3c.1.

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

Describe how feedback was obtained.

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

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

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.

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

This measure has been deemed a priority measure by NCQA and other entities, as illustrated by its use in programs such as the Annual State of Healthcare Quality and the Health Plan Rating.

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.

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.

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.

Performance on this measure has improved slightly for Medicaid and commercial health plans (see section 1b.2 for summary of data for commercial and Medicaid health plans). The largest improvement in performance scores is seen in the Medicaid product line. In 2019, a total of 273 commercial health plans and 169 Medicaid health plans across 50 states reported calendar year 2018 data on this measure. These data are nationally representative.

4b2. Unintended Consequences

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

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

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

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.

Yes

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

1932 : Diabetes Screening for People With Schizophrenia or Bipolar Disorder Who Are Using Antipsychotic Medications (SSD)

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

N/A

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?

No

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

The Metabolic Monitoring for Children and Adolescents on Antipsychotics measure assesses metabolic monitoring during the measurement year among children and adolescents who are prescribed antipsychotics. This measure is related to measure #1932 but addresses a different target population and measure focus. Measure #1932 assesses whether adults with schizophrenia or bipolar disorder who were prescribed antipsychotics are screened for diabetes. Similar to the Metabolic Monitoring for Children and Adolescents on Antipsychotics measure, this measure is specified for the health plan level

and uses administrative claims as the data source. The measures have different target populations but a similar measure focus. Measure #1932 focuses on adults 18 to 64 years of age who have schizophrenia or bipolar disorder and who are prescribed antipsychotics. The Metabolic Monitoring for Children and Adolescents on Antipsychotics measure includes all children and adolescents up to 17 years of age who are prescribed antipsychotics and does not focus on any specific conditions. Measure #1932 is focused on diabetes screening by receipt of a glucose test. While the Metabolic Monitoring for Children and Adolescents on Antipsychotics measure also includes assessing whether a glucose test was received, it additionally assesses whether a cholesterol test was received since the focus is not just diabetes screening. The two measures are aligned in the way glucose testing is identified and measured.

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, 202-955-3500-

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

Co.4 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-3500-

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.

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, CenCalHealth
Harmon Jordan, ScD, Westat
Virginia Raney, LCSW, Center for Medicaid and CHIP Services
Lynne Rothney-Kozlak, MPH, Rothney-KozlakConsulting, LLC
Laurie Spoll, Aetna

NATIONAL COLLABORATIVE FOR INNOVATION IN QUALITY MEASUREMENT (NCINQ) MEASUREMENT ADVISORY PANEL

Samantha Broderick, Cardinal Catalysts

Cathy Caldwell, MPH, Alabama Department of Public Health

Jim Chase, MHA, Network for Regional Healthcare Improvement

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

Kim Elliott, PhD, Health Services Advisory Group

Lynda Gargan, PhD, National Federation of Families for Children's Mental Health

David Harmon, Md, FAAP, Superior Health Plan

Jennifer Havens, MD, NYU Langone Health

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

Sheree Neese-Todd, MS, Rutgers University, Center for Health Services Research

Rob Penfold, PhD, Kaiser Permanente Washington Health Research Institute

Cheryl J. Roberts, JD, Virginia Department of Medical Assistance Services

Jeff Schiff, MD, MBA, Minnesota Department of Human Services

Julie Sonier, MPA, MN Community Measurement

Craig Thiele, MD, Health Management Associates

Jeb Weisman, PhD, Icahn School of Medicine at Mount Sinai

Judy Zerzan, MD, MPH, WA Health Care Authority

NCINQ CONSUMER PANEL

Joan Alker, MPhil, Georgetown Center for Children and Families

Roni Christopher, MEd, OTR/L, PCMH-CCE, The Greater Cincinnati Health Collaborative

Daniel Coury, MD, Nationwide Children's Hospital

Eileen Forlenza, Colorado Medical Home Initiative, Children and Youth with Special Health Care Needs Unit

Michaelle Gady, JD, Families USA

Janis Guerney, JD, Family Voices

Jocelyn Guyer, MPA, Georgetown Center for Children and Families

Catherine Hess, MSW, National Academy for State Health Policy

Carolyn Muller, RN, Montgomery County Health Department

Cindy Pellegrini, March of Dimes

Judith Shaw, EdD, MPH, RN, VCHIP

Stuart Spielman, JD, LLM, Autism Speaks

Michelle Sternthal, PhD, March of Dimes

NCINQ FOSTER CARE PANEL

Kamala Allen, MHS, Center for Health Care Strategies

Mary Applegate, MD, Ohio Department of Job and Family Services

Samantha Jo Broderick, Foster Care Alumni of America

Mary Greiner, MD, Cincinnati Children's Hospital Medical Center

David Harmon, MD, FAAP, Superior HealthPlan

Patricia Hunt, Magellan Health Services

Audrey LaFrenier, MSW, Parsons Child and Family Center

Bryan Samuels, MPP, Chapin Hall

Phil Scribano, DO, MSCE , The Children's Hospital of Philadelphia

Lesley Siegel, MD, State of Connecticut Department of Children and Families

Chauncey Strong, MSW, LGSW, Fairfax County Department of Family Services/Foster Care and Adoption

Janet (Jessie) Sullivan, MD, Hudson Health Plan

Nora Wells, MS, National Center for Family/Professional Partnerships

NCINQ MENTAL HEALTH PANEL

Francisca Azocar, PhD, Optum Health Behavioral Solutions

Frank Ghinassi, PhD, Western Psychiatric Institute and Clinic of UPMC Presbyterian Shadyside

Jennifer Havens, MD, NYU Langone Health

Danielle Laraque, MD, FAAP, Maimonides Infants and Children's Hospital of Brooklyn

NCINQ STATE PANEL

Mary Applegate, MD, Ohio Department of Job and Family Services

Susan Castellano, Minnesota Department of Human Services

Sharon Carte, MHS, State of West Virginia Children's Health Insurance Program

Catherine Hess, MSW, National Academy for State Health Policy

Michael Hogan, PhD, New York State office of Mental Health

Barbara Lantz, MN, RN, State of Washington Department of Social and Health Services, Medicaid Purchasing Administration

Judy Mohr Peterson, PhD, Oregon Health Authority

Tracy Plouck, MPA, Ohio Department of Mental Health

Gina Robinson, Colorado Department of Health Care Policy and Financing

Janet Stover, Illinois Association of Rehabilitation Facilities

Eric Trupin, PhD, University of Washington

NCINQ MEASURE DEVELOPMENT PARTNERS

Shahla Amin, MS, Rutgers University

Scott Bilder, PhD, Center for Health Services Research, Rutgers University

Stephen Crystal, PhD, Institute for Health, Health Care Policy and Aging Research, Rutgers University

Molly Finnerty, PhD, NY State Office of Mental Health

Emily Leckman-Westin, PhD, NY State Office of Mental Health

Sheree Neese-Todd, MA, Institute for Health, Health Care Policy and Aging Research, Rutgers University

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2014

Ad.3 Month and Year of most recent revision: 07, 2019

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? 06, 2020

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

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

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures, without modification, are encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Modifications to, and/or commercial use of, a measure requires the prior written consent of NCQA and is subject to a license at the discretion of NCQA. As used herein, "commercial use" refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.