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 #: 2801

Corresponding Measures:

De.2. Measure Title: Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics

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

De.3. Brief Description of Measure: Percentage of children and adolescents 1-17 years of age who had a new prescription for an antipsychotic medication, but no U.S. Food and Drug Administration primary indication for antipsychotics, and had documentation of psychosocial care as first-line treatment.

1b.1. Developer Rationale: This measure addresses use of first-line psychosocial care as one facet of safe and judicious use of antipsychotics in children and adolescents. Antipsychotic prescribing for youth has increased rapidly in recent decades. Although antipsychotic medications may serve as effective treatment for a narrowly defined set of psychiatric disorders in youth, they are often being prescribed for nonpsychotic conditions for which psychosocial interventions are considered first-line treatment. Thus, clinicians may be underutilizing safer first-line psychosocial interventions, and youth may be unnecessarily incurring the risks associated with antipsychotic medications and experiencing poorer mental and physical health outcomes.

S.4. Numerator Statement: Children and adolescents 1-17 years of age who had psychosocial care as first-line treatment prior to (or immediately following) a new prescription of an antipsychotic without a U.S. Food and Drug Administration primary indication for antipsychotic use.

S.6. Denominator Statement: Children and adolescents 1-17 years of age as of December 31 of the measurement year who had a new prescription of an antipsychotic medication for which they do not have a U.S. Food and Drug Administration primary indication for antipsychotics.

S.8. Denominator Exclusions: Exclude children and adolescents with a diagnosis of a condition for which antipsychotic medications have a U.S. Food and Drug Administration primary indication and are thus clinically appropriate: schizophrenia, schizoaffective disorder, bipolar disorder, other psychotic disorder, autism, or other developmental disorder.

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 New Measure

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**

Evidence Summary

The following clinical practice guidelines are the singular source of evidence supporting this process measure:

Practice Parameter for the Use of Atypical antipsychotic Medications in Children and Adolescents, July 2012 (2011 guidelines).

Recommendation 1: …education and psychotherapeutic interventions…." Should come be the ignition and during the treatment with atypical antipsychotics..

Recommendation 2: absent an FDA indication "or substantial evidence for effectiveness… consider other medications or psychosocial treatments before initiating antipsychotic therapy.

Graded as a "minimal standard/clinical standard," seemingly the highest grade the American Academy of Child and Adolescent reviews can assign.

Changes to evidence from last review

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

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

Exception to evidence

N/A

Questions for the Committee:

NONE

Guidance from the Evidence Algorithm

Box 5b→ Moderate rating

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

RATIONALE: Reasonable evidence, but just a single review.

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.

Commercial

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

2016|159|58%|11%|21%|43%|52%|58%|64%|71%|85%|12%

2017|187|54%|11%|21%|41%|46%|54%|61%|68%|82%|15%

2018|188|55%|11%|21%|41%|48%|55%|63%|70%|84%|14%

Medicaid

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

2016|134|60%|13%|4%|44%|54%|62%|68%|74%|86%|14%

2017|137|60%|13%|8%|46%|53%|61%|68%|73%|83%|15%

2018|139|58%|15%|5%|36%|53%|61%|67%|75%|89%|14%

Disparities

Specific data are not presented, except the Medicaid versus Commercial (they also mention Medicare, but do not report that) health insurance performance reports which do not really suggest substantial disparities and even show that Medicaid rates are generally a bit higher.

Disparities are noted in the literature:

1. MEPS shows that blacks and Latinos less likely to use outpatient MH Services (LeCook et al., 2003)
2. Unmet need for MH Services (Alegria et al., 2010)
3. Child welfare system involved has unmet need (Burns et al., 2004)
4. Antipsychotic use in foster care increased (Zito et al., 2008)
5. Medicaid and CHIP play significant role (Kataoka et al., 2001)

Questions for the Committee:

- None

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)

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."

- Practice guidelines provide evidence supporting the measure.
- The AACAP guideline from 2012 still appears to be the standard.
- evidence supports measure
- The evidence relates to this process measure. This measure is critical as it could prevent the unnecessary prescribing of antipsychotics to children and teens.
- The measure is directly related to the goal of decreasing off-label prescribing of antipsychotics in children ages 1-7.
- this is a maintenance measure. utilizes the practice parameter for use of atypical antipsychotic meds in children and adolescents from July 2012. Rec 1: education and psychotherapeutic interventions should come on the ignition during treatment with atypical antipsychotics. rec 2: absent an FDA indication or substantial evidence for effectiveness, CONSIDER other meds or psychosocial treatments before initiating antipsychotic therapy. a couple concerns i have with this measure. first: there may be situations where there is appropriate use of atypical antipsychotics without approved FDA indication before other psychosocial treatments may get established (situations where patient is at risk of higher level of care), or other diagnoses where it may be appropriate (DMDD did not exist as a diagnosis in 2012)... regardless, I'd rather have the focus be on measuring ANY CHILD (independent of diagnosis) who is on an atypical antipsychotic ALSO have other psychosocial care in place.
- No new studies that contradict current body of evidence
- Systematic review of the scientific literature examining the clinical validity of this measure not mentioned. Copy of paper provided to NQF staff. The main source of evidence is 2012 AACAP guidelines related to recommended care to AP Meds supporting risk of ap medses

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?

- Gaps in performance presented.
- Yes, performance data provided for 3 years, 2016,17,18 for both Commercial and Medicaid Plans. Medicaid performance was marginally better but overall room for improvement given the standard of practice has been in effect since 2011 and the initial NQF endorsement was 2016. was measure since
- yes

- Current data is provided and there is a performance gap. Blacks and Hispanics are less likely to seek mental health treatment.
- The measure demonstrated a meaningful performance gap across Medicaid and commercial plans.
- gap continues to exist.
- Performance has remained consistent across Medicaid and commercial health plans
- gap but little change 2016-2018

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

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

Validity

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

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

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**

NQF Staff Review

Testing data

- 2019 submission: HEDIS Medicaid and Commercial Health Plan Coverage
- 2015 submission: MAX (state); health plan data Medicaid and Commercial
- Level of analysis: Health Plan

Reliability methods

2019: Tested: 188 and 139, Commercial and Medicaid plans tested, respectively.

Mean denominator per plan: 138 and 494, respectively. In 2015 the MAX (8 states) mean was 832 and the Medicaid was 501 (n=17), and Commercial was 113 (n=13)

Social risk factors not analyzed: instead reporting by Medicaid ad Commercial

Reliability Results:

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

Product Line	Overall Reliability	Min	Percentiles					Max
			10th	25th	50th	75th	90th	
Commercial	0.797	0.469	0.529	0.610	0.710	0.829	0.907	0.957
Medicaid	0.980	0.757	0.866	0.929	0.975	0.988	0.992	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	.91
Medicaid Health Plans	.97	.77
Commercial Health Plans	.77	.53

Validity

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

Measure	**Pearson Correlation Coefficients**			
	Follow-Up After Hospitalization – 30 Days (6-17)	Follow-Up After Hospitalization – 7 Days (6-17)	Follow-Up After Emergency Department Visit – 30 Days (6-17)	Follow-Up After Emergency Department Visit – 7 Days (6-17)
First-line Psychosocial Care	0.45	0.45	0.52	0.43

**includes data submitted by 188 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**			
	Follow-Up After Hospitalization – 30 Days (6-17)	Follow-Up After Hospitalization – 7 Days (6-17)	Follow-Up After Emergency Department Visit – 30 Days (6-17)	Follow-Up After Emergency Department Visit – 7 Days (6-17)
First-line Psychosocial Care	0.46	0.53	0.64	0.62

***includes data submitted by 139 Medicaid plans to HEDIS for these measures for measurement year 2018*

Note: All correlations are significant at p<.0001

2015 results:

Measure	Pearson Correlation Coefficients		
	First-Line Psychosocial Care	Follow-Up Visit	Metabolic Screening
First-Line Psychosocial Care	1	0.59 (p=.03)	0.18 (p=.55)
Follow-Up Visit		1	0.06
Metabolic Screening			1

Face validity: 80% public comment support after 2013-4 NCQA expert panels developed the (5 panels worked together)

Interpretation: Pearson Coefficients: .3 to .59 are considered moderate associations

Questions for the Committee regarding reliability:

- Any concern about the commercial plan reliability given that the 25th percentile has Adams R scores well below the heuristic threshold?

Questions for the Committee regarding validity:

- Any concern about the comparators chosen as external standards to this measure of proper sequencing in prescribing antipsychotics? Are the comparators too tangential to be validators? Is the metabolic screening result invalidating?

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

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

Evaluation A: Scientific Acceptability

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 2801

Measure Title: Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics Use of First-Line Psychosocial Care 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.**

   No concerns from NQF staff, but Committee should agree to the details of the indications and pharmaceuticals included/excluded.

   Drug list: 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

   Psychotherapeutic combinations / Fluoxetine-olanzapine; Perphenazine-amitriptyline

   Psychosocial therapies list: 458 codes in that value set

   Exclusions as FDA approved indications: Schizophrenia Value Set; Bipolar Disorder Value Set; Other Psychotic and Developmental Disorders Value Set

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   (see description above)

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

   Submission document: Testing attachment, section 2a2.3

   (see description above)

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

(see notations above and below)

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

    Submission document: Testing attachment, section 2b2.

Two points are made in this section for consideration by the Committee. First is this statement from the developer's application.

> "The rates of antipsychotics usage amongst those with an FDA-indication–the excluded population–also match what the literature has indicated, which were found to be approximately 30% (Sohn et al., 2016)."

It suggests one point of concern regarding exclusions, they are quite comment. At the same time, the below table shows that the performance rates with and without exclusions are comparable. Finally, the above indicator (that only 30% of prescriptions in youths are FDA-indicated) obviously shows that off-label use is very prevalent. This point, perhaps, increases the importance of this particular measure.

2015 submission data: Exclusion for Diagnosis during Measurement Year that has FDA Indication for Antipsychotics

	Age 0-5			Age 6-11			Age 12-17		
	# with exclusion diagnosis	Rate with exclusion	Rate without exclusion	# with exclusion diagnosis	Rate with exclusion	Rate without exclusion	# with exclusion diagnosis	Rate with exclusion	Rate without exclusion

Plan 1	0	25.0%	25.0%	53	63.2%	64.9%	102	73.6%	73.8%
Plan 2	3	75.0%	42.9%	8	76.9%	70.6%	23	77.8%	79.1%
Plan 3	4	33.3%	23.1%	26	35.2%	42.1%	59	54.2%	55.9%
Plan 4	0	NA	NA	4	76.5%	76.2%	13	64.0%	65.8%
Plan 5	0	NA	NA	12	75.0%	66.7%	23	65.8%	69.6%
Plan 6	2	25.0%	50.0%	9	59.4%	58.5%	21	51.2%	55.2%
Plan 7	0	100.0%	100.0%	7	64.3%	71.4%	23	57.4%	68.6%
Plan 8	0	0.0%	0.0%	4	61.5%	70.6%	20	67.8%	72.2%
Plan 9	0	50.0%	50.0%	6	18.2%	35.3%	15	50.0%	52.7%
Plan 10	0	0.0%	0.0%	10	66.7%	60.0%	10	69.8%	71.4%
Plan 11	1	60.0%	66.7%	4	81.8%	66.7%	17	70.7%	70.7%
Plan 12	1	33.3%	25.0%	4	54.5%	66.7%	18	57.1%	60.9%
Plan 13	1	NA	100.0%	5	71.4%	58.3%	10	50.0%	47.5%
Total	**12**	**40.0%**	**38.3%**	**152**	**57.7%**	**59.5%**	**354**	**63.7**	**66.0%**

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

    Submission document: Testing attachment, section 2b4.

    No concerns, they are evident:

2019 Submission:

Table 5. Variation in Performance for Commercial and Medicaid health plans, 2018.

Plan Type	N	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Commercial	188	55.1%	10.6%	41.0%	48.3%	54.8%	62.6%	69.6%	14.3%	<0.0001
Medicaid	139	57.6%	15.5%	36.4%	52.7%	60.6%	66.6%	75.0%	13.9%	<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.

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.

NONE

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**

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

(see above)

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

Submission document: Testing attachment, section 2b2.3

(see above)

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?

- No concerns.
- No concerns.
- no issues
- From the evidence provided, this measure can be consistently implemented.
- According to the Submission, there is a rigorous audit process to ensure the eligible population, denominator, and numerator events for each measure are correctly identified and reported. Further, the audit process appears to be designed to verify primary data sources used to populate measures and ensure specifications are correctly implemented.
- how is "immediately following" defined? in the numerator (psychosocial care as first line treatment prior to (or immediately following) a new prescription without FDA indication)
- All data elements are in defined fields in electronic claims
- specifications are clearly described, purposively excludes children and adolescents for whom first-line ap meds may be clinically appropriate.

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

- No.

- No.
- no
- No--this measure has moderate reliability.
- No
- -
- No concerns
- applied NCQA's approach for health plan level data

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

- no
- No.
- no
- No--the measure has face and construct validity.
- No
- no concerns
- No concerns
- applied usual approach, Examined convergent validity using correlation with the Follow-up Measure after Hospitalization for MI and ED for MI. correlations ranged form .43-.62, moderate

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
- None
- no concerns
- I do not believe there are any major threats to validity.
- There appears to be minimal threats to validity. The measure uses an annual audit process to identify missing data in a structured way. Further, NCQA-certified auditors use standard audit methodologies to assess whether data sources are missing data.
- no concerns
- No concerns
- difficult to assess a "meaningful difference" based on data presented

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?

- yes
- The exclusions are appropriate and there was no risk adjustment. I agree that the data illustrating only 30% of the prescriptions were indicated speaks to the need to continue to measure.
- no need to risk adjust
- No patients are inappropriately excluded.
- The exclusions are consistent with the evidence and there no inappropriate exclusions. The analysis indicates acceptable results.
- DMDD? mood d/o NOS? treatment resistant depression? there may be other situations where prescribing may be appropriate (DMDD, mood disorder NOS, treatment resistant depression, situations where patient is at risk for higher level of care
- No risk adjustment; no concerns about exclusions
- appropriately excludes children/teens with dx for which ap meds are first line. What is the remaining n after this exclusion?

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 generated as part of provision of care
- All electronic
- Subject to well-established NCQA compliance audits
- Real-time feedback from measure user system (thousands on over 100 measures)

Questions for the Committee:

- NONE

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?

- Data obtained from EHR, no concerns.
- No concerns.
- feasible
- The data is generated and available through electronic sources.
- I was unable to discren from the Submission which data elements are available in electronic form.

- no concerns
- All data elements are in electronic claims
- to be used at health plan level. If adapted for Medicaid claims the problem of lack of codes for evidence-based psychosocial treatments remains

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

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.

Medicaid Child Core Set: This measure is used in the Child Core Set administered by the Centers for Medicare and Medicaid Services (CMS). The Child Core Set intends to evaluate the quality of care and outcomes among children in Medicaid and Children's Health Insurance Program (CHIP). The measure set is reported by over 25 states annually and performance is available publicly at Medicaid.gov."

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

Feedback on the measure by those being measured or others: The measure was said to have been developed by 5 TEPs working together, and the measure then received 80 percent approval ratings from broader stakeholder audiences.

Additional Feedback: A general narrative was given in the application describing that results from NCQA are regularly published and that the regularly collect information from the plans being measured absent the expression of any significant concerns about this particular metric.

***Questions for the Committee**:*

- NONE

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 In section 4b1, the developers note that over the first 3 years of the measures' life no improvement was observed. As a explanation for this flat trajectory, they further speculate that it will take more time for measured system to "bridge the fragmentation gaps" necessary for improvement in this metric to be realized.

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

Unexpected findings (positive or negative) during implementation NONE

Potential harms NONE

Additional Feedback: NONE

***Questions for the Committee**:*

- Concern about the flat trajectory for performance on this measure over the first three years of its existence?

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?

- yes.
- Is publicly reported and used in multiple accountability applications, both Commercial and Medicaid. Feedback has been considered.
- It is not clear that feedback was received by NCQA or considered because there was no detail on this in the application.
- The data is publicly reported.
- As this is health plan data it is likely accessible through accreditation processes such as URAC. Further, in the case of Medicaid plans, it is likely avaiable through state oversight.
- currently in use; medicaid core set.
- Currently being used in publicly reported and accountability and benchmarking applications with feedback through NCQA
- reported publicly by NCQA

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.

- no concerns
- There is continuing rationale to continue to measure and use these performance results. No unintended consequences.
- may be specific but not sensitive because claims data often undercode serious MH diagnoses such as schizophrenia and bipolar
- There is good benefit to measurement. There are no harms.
- This measure is important for the goal of reducing inappropriate prescribing of antipsychotics in children. The only potential harm is children being referred to psychosocial interventions (and providers) that lack fidelity to evidence-based practice and methods.
- performance has remained consistent in the past several years -? improving care...
- Plans report no significant barriers to implementing the measure even though measure requires coordination across settings and providers
- relatively little change in performance across the 3 years reported

Criterion 5: Related and Competing Measures

Related or competing measures
None reported by developer, and none found by NQF staff
Harmonization
N/A

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?

- no related/competing measures
- N/A
- not that I know of
- There are no competing measures.
- None that I am aware of.
- none
- No related or competing measures

Public and Member Comments

No comments and member support/non-support submitted as of 01/23/2020.

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

APP_Evidence_Form_-2801-.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*)**:** 2801

Measure Title: Use of First-Line Psychosocial Care 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: Click here to enter a date

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: Use of psychosocial care used as a first-line therapy for children newly started on antipsychotics without an FDA primary indication

☐ 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.

Child does NOT have a primary indication for antipsychotic use >>> Health care provider utilizes psychosocial care intervention >>> Child avoids unnecessary antipsychotic use >>> Child avoids adverse

side effects associated with antipsychotic medications >>> Child experiences improvement in mental and physical outcomes (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 Education.

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,	Recommendation 1: • "Prior to the initiation of and during treatment with an AAA, the general guidelines that pertain to the prescription of psychotropic medications should be followed… *including*

summarize the conclusions from the SR.	*education and psychotherapeutic interventions for the treatment and monitoring of improvement.*"
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	Grade of Recommendation: Clinical Guideline

recommendation grading system	• 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?	The American Academy of Child and Adolescent Psychiatry (AACAP) guideline recommends use of psychosocial services prior to antipsychotics initiation, particularly in the absence of an FDA indication. These recommendations are based on established metabolic and other health risks of antipsychotics as well as evidence of efficacy of psychosocial treatments.
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: • The AACAP-AAA guideline is rated a ***Clinical Standard***, indicating it is based on rigorous/substantial empirical evidence and/or overwhelming clinical consensus. AACAP includes several condition-specific guidelines around use of psychosocial services; we focus on the general AACAP-AAA antipsychotics guideline here and describe the body of evidence for each relevant recommendation below. • When developing their guideline, 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. AACAP did not provide a breakdown of specific numbers of each publication type. We have identified where there are certain publication types available to support each guideline. • This recommendation is based on a literature review conducted by a medical professional society on the established metabolic impacts of antipsychotics and other health risks and evidence of efficacy of psychosocial treatments. The literature review contained a total of 147 publications that included clinical trials,

	meta-analysis, practice guidelines, RCTs, systematic literature reviews, and case reports and series. • American Academy of Child and Adolescent Psychiatry. Practice parameter on the use of psychotropic medications in children and adolescents. *J Am Acad Child Adolesc Psychiatry.* 2009;48:961-973. 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 use of first-line psychosocial care for children and adolescents on antipsychotics is moderate to high.
Estimates of benefit and consistency across studies	The AACAP-AAA review did not include an exact estimate of benefits of psychosocial care. However, the evidence has established that antipsychotic use is associated with adverse short-term metabolic and other side effects in youth and with negative long-term health outcomes throughout the lifespan.
What harms were identified?	The AACAP review did not examine the potential harms of treating children with psychosocial care prior to initiating antipsychotics. However, the harms of unnecessary antipsychotic use in kids has been well established (Andrade et al. 2011; Bobo et al., 2013; Correll, 2008; Correll et al., 2009; Crystal et al., 2009; Daniels, 2006; Lean and Pajonk, 2003; Srinivasan et al. 2002). Citations: Andrade, S.E., J.C. Lo, D. Roblin, et al. December 2011. Antipsychotic medication use among children and risk of diabetes mellitus. *Pediatrics*. 128(6):1135–41. Bobo, W.V., W.O. Cooper, C.M. Stein, et al. October 1, 2013. Antipsychotics and the risk of type 2 diabetes mellitus in children and youth. *JAMA Psychiatry.* 70(10):1067–75. Correll, C.U. 2008. Antipsychotic use in children and adolescents: minimizing adverse effects to maximize outcomes. *FOCUS: The Journal of Lifelong Learning in Psychiatry.* 6(3):368–78. Correll, C. U., Manu, P., Olshanskiy, V., Napolitano, B., Kane, J. M., & Malhotra, A. K. 2009. Cardiometabolic risk of second-generation antipsychotic medications during first-time use in children and adolescents. *Journal of the American Medical Association*. 302(16):1765-1773. Crystal, S., M. Olfson, C. Huang, H. Pincus and T. Gerhard. 2009. Broadened use of atypical antipsychotics: Safety, effectiveness, and policy challenges. *Health Affairs.* 28:w770–81.

	Daniels, S.R. 2006. The consequences of childhood overweight and obesity. *The future of children.* 16(1):47–67. Lean, M.E., and F.G. Pajonk. 2003. Patients on Atypical Antipsychotic Drugs Another high-risk group for type 2 diabetes. *Diabetes Care.* 26(5), 1597–605. Srinivasan, S. R., Myers, L., & Berenson, G. S. 2002. Predictability of childhood adiposity and Insulin for developing insulin resistance syndrome (syndrome X) in young adulthood the Bogalusa heart study. *Diabetes.* 51(1):204-209.
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 Treatments.

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 2: • "*In the absence of specific FDA indications or substantial evidence for effectiveness, physicians should consider other medication or psychosocial treatments before initiating antipsychotic treatment*."
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?	The American Academy of Child and Adolescent Psychiatry (AACAP) guideline recommends use of psychosocial services prior to antipsychotics initiation, particularly in the absence of an FDA indication. These recommendations are based on established metabolic

	and other health risks of antipsychotics as well as evidence of efficacy of psychosocial treatments.
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: • The AACAP-AAA guideline is rated a ***Clinical Standard***, indicating it is based on rigorous/substantial empirical evidence and/or overwhelming clinical consensus. AACAP includes several condition-specific guidelines around use of psychosocial services; we focus on the general AACAP-AAA antipsychotics guideline here and describe the body of evidence for each relevant recommendation below. • When developing their guideline, 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. AACAP did not provide a breakdown of specific numbers of each publication type. We have identified where there are certain publication types available to support each guideline. • This recommendation is based on literature regarding the use of antipsychotics in specific clinical populations and the current FDA indications, which include only schizophrenia, bipolar disorder, tic disorders and specific symptoms of autistic disorder. In the absence of substantial empirical support for antipsychotics for other specific problems or specific FDA indications, AACAP recommends health care providers implement other pharmacological or psychosocial treatment modalities with more established efficacy and safety profiles prior to the onset of antipsychotics use. 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 use of first-line psychosocial care for children and adolescents on antipsychotics is moderate to high.
Estimates of benefit and consistency across studies	The AACAP-AAA review did not include an exact estimate of benefits of psychosocial care. However, the evidence has established that antipsychotic use is associated with adverse short-term metabolic and

	other side effects in youth and with negative long-term health outcomes throughout the lifespan.
What harms were identified?	The AACAP review did not examine the potential harms of treating children with psychosocial care prior to initiating antipsychotics. However, the harms of unnecessary antipsychotic use in kids has been well established (Andrade et al. 2011; Bobo et al., 2013; Correll, 2008; Correll et al., 2009; Crystal et al., 2009; Daniels, 2006; Lean and Pajonk, 2003; Srinivasan et al. 2002). Citations: Andrade, S.E., J.C. Lo, D. Roblin, et al. December 2011. Antipsychotic medication use among children and risk of diabetes mellitus. *Pediatrics*. 128(6):1135–41. Bobo, W.V., W.O. Cooper, C.M. Stein, et al. October 1, 2013. Antipsychotics and the risk of type 2 diabetes mellitus in children and youth. *JAMA Psychiatry.* 70(10):1067–75. Correll, C.U. 2008. Antipsychotic use in children and adolescents: minimizing adverse effects to maximize outcomes. *FOCUS: The Journal of Lifelong Learning in Psychiatry.* 6(3):368–78. Correll, C. U., Manu, P., Olshanskiy, V., Napolitano, B., Kane, J. M., & Malhotra, A. K. 2009. Cardiometabolic risk of second-generation antipsychotic medications during first-time use in children and adolescents. *Journal of the American Medical Association*. 302(16):1765-1773. Crystal, S., M. Olfson, C. Huang, H. Pincus and T. Gerhard. 2009. Broadened use of atypical antipsychotics: Safety, effectiveness, and policy challenges. *Health Affairs.* 28:w770–81. Daniels, S.R. 2006. The consequences of childhood overweight and obesity. *The future of children.* 16(1):47–67. Lean, M.E., and F.G. Pajonk. 2003. Patients on Atypical Antipsychotic Drugs Another high-risk group for type 2 diabetes. *Diabetes Care.* 26(5), 1597–605. Srinivasan, S. R., Myers, L., & Berenson, G. S. 2002. Predictability of childhood adiposity and Insulin for developing insulin resistance syndrome (syndrome X) in young adulthood the Bogalusa heart study. *Diabetes.* 51(1):204-209.
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 use of first-line psychosocial care as one facet of safe and judicious use of antipsychotics in children and adolescents. Antipsychotic prescribing for youth has increased rapidly in recent decades. Although antipsychotic medications may serve as effective treatment for a narrowly defined set of psychiatric disorders in youth, they are often being prescribed for nonpsychotic conditions for which psychosocial interventions are considered first-line treatment. Thus, clinicians may be underutilizing safer first-line psychosocial interventions, and youth may be unnecessarily incurring the risks associated with antipsychotic medications and experiencing poorer mental and physical health outcomes.

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 is 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 psychosocial care as first-line treatment prior to (or immediately following) a new prescription of an antipsychotic.

Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics

N = Number of Health Plans

YEAR = Measurement Year

Commercial

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

2016|159|58%|11%|21%|43%|52%|58%|64%|71%|85%|12%

2017|187|54%|11%|21%|41%|46%|54%|61%|68%|82%|15%

2018|188|55%|11%|21%|41%|48%|55%|63%|70%|84%|14%

Medicaid

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

2016|134|60%|13%|4%|44%|54%|62%|68%|74%|86%|14%

2017|137|60%|13%|8%|46%|53%|61%|68%|73%|83%|15%

2018|139|58%|15%|5%|36%|53%|61%|67%|75%|89%|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

Research using the Medical Expenditure Panel Survey shows that black and Latino youth ages 5-21 were significantly less likely to access outpatient mental health care (Le Cook et al., 2013). This finding is consistent with research suggesting that minority youth may have both higher unmet needs for mental health care and receive lower-quality care compared to white American youth (Alegria et al., 2010). Data also suggest that youth involved in the child welfare system, particularly those ages ten and younger, may have significant unmet mental health needs (Burns et al., 2004). Analysis of Medicaid data shows that youth in foster care are more likely to be prescribed antipsychotic medications compared to those not in foster care (Zito et al., 2008). These trends suggest that access to psychosocial interventions for minority and foster care youth prescribed antipsychotics may be of particular importance.

Research also demonstrates that children without health insurance have higher rates of unmet needs for mental health care compared to those with public insurance, suggesting that Medicaid and the Children's Health Insurance Program may play an important role in promoting access to care (Kataoka et al., 2001). In addition, the rate of increase in use of antipsychotics is higher for children and adolescents with public

insurance than commercial insurance, suggesting this measure may particularly help improve the quality of mental health care for children with public insurance. It is unclear what factors are associated with lack of access to psychotherapy for this population.

Alegria M, Vallas M, Pumariega AJ. Racial and ethnic disparities in pediatric mental health. Child Adolesc Psychiatr Clin N Am. Oct 2010; 19(4):759-774.

Burns BJ, Phillips SD, Wagner HR, et al. Mental health need and access to mental health services by youths involved with child welfare: a national survey. J Am Acad Child Adolesc Psychiatry. Aug 2004; 43(8):960-970.

Kataoka SH, Zhang L, Wells KB. Unmet need for mental health care among U.S. children: variation by ethnicity and insurance status. Am J Psychiatry. Sep 2002; 159(9):1548-1555.

Le Cook B, Barry CL, Busch SH. Racial/ethnic disparity trends in children's mental health care access and expenditures from 2002 to 2007. Health Serv Res. Feb 2013; 48(1):129-149.

Zito JM, Safer DJ, Sai D, et al. Psychotropic medication patterns among youth in foster care. Pediatrics. Jan 2008; 121(1):e157-163.

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):*

Access to Care, Safety, Safety : Medication

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:** 2801_APP_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. For maintenance of endorsement, 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.

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

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

Children and adolescents 1-17 years of age who had psychosocial care as first-line treatment prior to (or immediately following) a new prescription of an antipsychotic without a U.S. Food and Drug Administration primary indication for antipsychotic use.

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

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

The numerator is reported using administrative data and includes children and adolescents who had documentation of psychosocial care (Psychosocial Care Value Set) in the 121-day period spanning 90 days prior to the IPSD through 30 days after the IPSD during the measurement year (January 1 – December 1).

The IPSD is earliest prescription dispensing date for an antipsychotic medication where the date is in the Intake Period and there is a Negative Medication History 120 days (4 months) prior to the IPSD when the member had no antipsychotic medications dispensed for either new or refill prescriptions.

See attachment for all value sets reference above (S.2b).

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

Children and adolescents 1-17 years of age as of December 31 of the measurement year who had a new prescription of an antipsychotic medication for which they do not have a U.S. Food and Drug Administration primary indication for antipsychotics.

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 year as of December 31 of the measurement year who had a new prescription for an antipsychotic medication (Table APP-A) during the Intake Period. Details to identify the eligible population are below.

STEP 1: Identify all patients in the specified age range who were dispensed an antipsychotic medication (Table APP-A) during the Intake Period.

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

STEP 3: Calculate continuous enrollment. Members must be continuously enrolled for 120 days (4 months) prior to the IPSD through 30 days after the IPSD.

TABLE APP-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

Psychotherapeutic combinations / Fluoxetine-olanzapine; Perphenazine-amitriptyline

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

Exclude children and adolescents with a diagnosis of a condition for which antipsychotic medications have a U.S. Food and Drug Administration primary indication and are thus clinically appropriate: schizophrenia, schizoaffective disorder, bipolar disorder, other psychotic disorder, autism, or other developmental disorder.

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 children and adolescents for whom first-line antipsychotic medications may be clinically appropriate. Any of the following during the measurement year meet criteria:

- At least one acute inpatient encounter with a diagnosis of schizophrenia, schizoaffective disorder, bipolar disorder, other psychotic disorder, autism, or other developmental disorder during the measurement year. Any of the following code combinations meet criteria:

-- BH Stand Alone Acute Inpatient Value Set with (Schizophrenia Value Set; Bipolar Disorder Value Set; Other Psychotic and Developmental Disorders Value Set).

-- Visit Setting Unspecified Value Set with Acute Inpatient POS Value Set with (Schizophrenia Value Set; Bipolar Value Set; Other Psychotic and Developmental Disorders Value Set).

- At least two visits in an outpatient, intensive outpatient or partial hospitalization setting, on different dates of service, with a diagnosis of schizophrenia, schizoaffective disorder, bipolar disorder, other psychotic disorder, autism, or other developmental disorder during the measurement year. Any of the following code combinations with (Schizophrenia Value Set; Bipolar Disorder Value Set; Other Psychotic and Developmental Disorders Value Set) meet criteria:

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

-- An outpatient visit (BH Outpatient Value Set).

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

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

-- Electroconvulsive therapy (Electroconvulsive Therapy Value Set).

-- An observation visit (Observation Value Set).

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

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 as of December 31 of the measurement year.
- EVENT/DIAGNOSIS: Identify the number of children and adolescents who were newly dispensed an antipsychotic medication during the intake period. SEE S.7 for the list of antipsychotic medications.

STEP 2: Exclude patients who meet the exclusion criteria. SEE S.8 and S.9 for denominator exclusion criteria and details.

STEP 3: Determine the numerator by identifying the number of children and adolescents in the eligible population who had documentation of psychosocial care in the 121-day period spanning 90 days prior through 30 days after the new prescription of an antipsychotic.

STEP 4: Calculate the rate by dividing the numerator (Step 3) by the denominator (after exclusions) (Step 2).

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)*

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

APP_Testing_Form_-2801-_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*)**:** 2801

Measure Title: Use of First-Line Psychosocial Care 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 https://www.medicaid.gov/medicaid/data-and-systems/macbis/medicaid-and-chip-research-files/max/index.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, 2010 Medicaid health plan data for 17 plans, and 2012 commercial health plan data for 73 plans.

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: *Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics* is a Healthcare Effectiveness Data and Information Set (HEDIS) health-plan level measure that assesses whether youth who had a new prescription for an antipsychotic without a primary indication, had documentation of psychosocial care as first-line treatment. We tested the measure in health plans, 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 188 commercial plans and 139 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: As part of the Pediatric Quality Measures Program (PQMP), NCINQ had access to the Medicaid Analytic eXtract (MAX) for conducting state analyses. In addition, NCINQ was able to test this

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

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 19 commercial health plans nationwide

These administrative data sources included claims for all 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 Mathematic Policy Research reports that suggested that they provided adequate encounter/managed care data (Byrd & Dodd, 2012; Byrd & Dodd, 2013). Of these 11 states, three were excluded in the testing for this measure due to lack of completeness of data.

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. 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 Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics, 2018.

Product Line	Number of Plans	Mean Denominator Size/Plan
Commercial	188	138
Medicaid	139	494

2015 Submission: We tested a set of several measures related to antipsychotic use in 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 14,598 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 8,525, and across 13 commercial plans the total was 1,472.

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 Eight States (MAX) (2008)

Mean	832
Minimum	269
25th	371
Median	1,350
75th	1,990
Maximum	3,376

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

Mean	501
Minimum	53
25th	133
Median	426
75th	749
Maximum	1,384

Denominator Size Distribution Across 13* Commercial Health Plans Nationwide (2012)

Mean	113
Minimum	37
25th	48
Median	70
75th	112
Maximum	387

*Of the 19 plans included in the testing of this measure, 13 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 health plan 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 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.

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 testing the impact of exclusions, the commercial health plan data sample was used.

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 no 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 and the distribution of individual plan reliability.

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

Product Line	Overall Reliability	Min	Percentiles					Max
			10th	25th	50th	75th	90th	
Commercial	0.797	0.469	0.529	0.610	0.710	0.829	0.907	0.957
Medicaid	0.980	0.757	0.866	0.929	0.975	0.988	0.992	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	.91

Medicaid Health Plans	.97	.77
Commercial Health Plans	.77	.53

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 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 Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics measure by exploring whether it was correlated with other similar measures of quality hypothesized which are listed below.

- Follow-Up After Hospitalization for Mental Illness – 30 days (ages 6 – 17): The percentage of discharges for patients 6 to 17 years of age, who were hospitalized for treatment of selected mental illness or intentional self-harm diagnoses, for which the patient received follow-up with a mental health practitioner within 30 days after discharge.
- Follow-Up After Hospitalization for Mental Illness – 7 days (ages 6 – 17): The percentage of discharges for patients 6 to 17 years of age, who were hospitalized for treatment of selected mental illness or intentional self-harm diagnoses, for which the patient received follow-up with a mental health practitioner within 7 days after discharge.
- Follow-Up After Emergency Department Visit for Mental Illness – 30 days (ages 6 – 17): The percentage of emergency department (ED) visits for patients 6 to 17 years of age with a principal diagnosis of mental illness or intentional self-harm for which the patient received follow-up for mental illness within 30 days of the ED visit (31 total days).
- Follow-Up After Emergency Department Visit for Mental Illness – 7 days (ages 6 – 17): The percentage of emergency department (ED) visits for patients 6 to 17 years of age with a principal diagnosis of mental illness or intentional self-harm for which the patient received follow-up for mental illness within 7 days of the ED visit (8 total days).

These measures were chosen for construct validity testing because these measures also capture the coordination of care and services for patients presenting with mental illness. These measures evaluate timely

access to appropriate mental health services following a high-intensity event. Patients should similarly have access to appropriate psychosocial care prior to initiation of an antipsychotic regimen. We hypothesized that organizations that perform well on the HEDIS Follow-up measures should perform well on the Psychosocial Care HEDIS 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 than 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 two related measures: *Metabolic Screening for Children and Adolescents on Antipsychotics* and *Follow-Up Visit for Children and Adolescents on Antipsychotics* measures. 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, and 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. 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 call the HEDIS measure life cycle. This measure has undergone the following steps associated with that cycle.

Step 1: NCQA staff identifies areas of interest of 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 includes 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 Update:
Construct Validity
The results from construct validity testing of the health plan level measure are presented by product line in Tables 3a and 3b below.

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

Measure	Pearson Correlation Coefficients			
	Follow-Up After Hospitalization – 30 Days (6-17)	Follow-Up After Hospitalization – 7 Days (6-17)	Follow-Up After Emergency Department Visit – 30 Days (6-17)	Follow-Up After Emergency Department Visit – 7 Days (6-17)
First-line Psychosocial Care	0.45	0.45	0.52	0.43

**includes data submitted by 188 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			
	Follow-Up After Hospitalization – 30 Days (6-17)	Follow-Up After Hospitalization – 7 Days (6-17)	Follow-Up After Emergency Department Visit – 30 Days (6-17)	Follow-Up After Emergency Department Visit – 7 Days (6-17)
First-line Psychosocial Care	0.46	0.53	0.64	0.62

***includes data submitted by 139 Medicaid plans to HEDIS for these measures for measurement year 2018*
Note: All correlations are significant at p<.0001

2015 Submission:
Construct Validity

Correlations

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 moderate positive correlation between the *Follow-Up Visit* and *Psychosocial Care* measures (r=0.59, p=0.03) and very slight positive correlation between the *Metabolic Screening* and *Psychosocial Care* measures (r=0.18, p=.55).

Measure	**Pearson Correlation Coefficients**		
	First-Line Psychosocial Care	Follow-Up Visit	Metabolic Screening
First-Line Psychosocial Care	1	0.59	0.18
Follow-Up Visit		1	0.06
Metabolic Screening			1

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

State	**First-Line Psychosocial Care**	**Follow-Up Visit**	**Metabolic Screening**
1	**36.7**	**60.2**	**2.6**
2	**35.8**	**68.4**	4.5
3	60.3	75.0	5.5
4	48.9	71.2	3.8
5	45.0	74.9	**0.4**
6	64.1	76.4	4.8
7	41.5	69.0	6.3
8	N/A*	N/A*	5.3
9	N/A*	N/A*	**10.7**
10	**53.3**	**81.3**	8.3
11	N/A*	**78.8**	**14.0**
Mean	**48.2**	**72.8**	**6.0**

*State was excluded from analysis due to incomplete data

Medicaid Health Plan Performance Rankings for One State

Plan	**First-Line Psychosocial Care**	**Follow-Up Visit**	**Metabolic Screening**
3	41.7	**71.0**	**0.2**
9	48.6	81.8	**4.9**
6	30.1	83.5	12.3
17	**26.4**	**86.7**	14.8
2	**27.4**	80.5	**15.4**
8	43.5	81.1	12.6
4	46.9	78.7	9.3
5	42.4	80.0	10.6

1	51.6	82.1	12.8
11	43.8	74.4	6.1
16	56.6	78.8	10.6
15	28.0	80.9	10.8
12	43.3	77.2	13.3
13	30.7	**70.4**	**17.8**
7	**67.7**	85.3	5.1
14	64.3	**98.7**	7.1
10	**67.0**	78.9	10.6
Mean	**44.7**	**80.6**	**10.3**

Face Validity

Step 1: This measure was developed to address the need for first-line psychosocial care for those youth started on antipsychotics who do not have a primary indication for an antipsychotic. NCQA and five expert panels worked together in 2013 and 2014 to identify the most appropriate method for assessing first-line psychosocial care among this patient population. Across the multiple expert panels that reviewed this measure, all panels concluded this measure was specified to assess the use of psychosocial care as first-line treatment for children without a primary indication for antipsychotics.

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 73 comments received, the vast majority (80 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 result 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 follow-up measures for individuals with mental illness are correlated to performance on the Use of First-Line Psychosocial Care 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 commercial plans that performed well on providing follow-up visits for those newly prescribed antipsychotics also performed well on providing first-line psychosocial care to those newly on antipsychotics. There was also a very slight positive correlation between the *Psychosocial Care* measure and *Metabolic Screening* measures, indicating that plans that perform well on providing first-line psychosocial care also perform well on providing baseline metabolic screening for those newly prescribed 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 these measures suggest the measures are assessing a common dimension of quality.

Face Validity

The expert panels consulted 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*

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*)

2019 Submission: This measure assesses the use of psychosocial care as a safer first-line alternative to antipsychotics for individuals without a primary indication for antipsychotics. Thus, the measure excludes youth with conditions for which there is a U.S. Food and Drug Administration indication for antipsychotics (schizophrenia, bipolar disorder, psychotic disorder, autism, tic disorders). The intent of these exclusions is to remove a population that is characteristically different and would not necessarily be recommended for psychosocial care first. We analyzed the distribution of the rate of exclusions across product lines and age stratifications in the measure using HEDIS data (detailed below) to assess the percentage of children removed due to the exclusion and if this is consistent with what we would expect to see based on what we know from the literature on the prevalence of these conditions.

2015 Submission: The measure excludes youth with conditions for which there is a U.S. Food and Drug Administration indication for antipsychotics (schizophrenia, bipolar disorder, psychotic disorder, autism, tic disorders). We tested the impact of exclusions using the commercial health plan data. The aim of testing exclusions in the field test data was to determine how common exclusions are in the eligible patient population and the impact of these exclusions on denominator sizes and performance rates. Our results (detailed below) show differences in performance rates with and without exclusions.

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*)

2019 Submission:

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

Plan Type	STRAT	N	Avg.	10th	25th	50th	75th	95th	IQR
Commercial*	6-11 years	52	30.3%	19.5%	25.5%	30.2%	34.7%	44.3%	9.2%
	12-17 years	169	29.1%	22.0%	25.8%	29.6%	32.8%	38.9%	7.0%
	total	188	29.8%	23.0%	26.3%	30.1%	33.3%	39.1%	7.0%
Medicaid	1-5 years	17	26.8%	16.0%	18.9%	26.2%	30.6%	50.0%	11.7%
	6-11 years	113	24.2%	15.9%	18.9%	23.0%	28.3%	40.1%	9.4%

	12-17 years	132	29.2%	20.2%	24.8%	27.8%	33.7%	44.3%	8.9%
	total	139	27.4%	18.5%	22.8%	26.3%	31.4%	41.1%	8.6%

STRAT = Age stratification with total indicating a sum of stratifications
N = Number of plans reporting
IQR = Interquartile range
**Note: 1-5 years stratification analysis for the Commercial product line was not run since Commercial plan denominator sizes were too small.*

2015 Submission: On average 25% of children age 0-5 with a new start of an antipsychotic met the exclusion criteria for having a primary indication for antipsychotic use (i.e., schizophrenia, bipolar disorder, psychotic disorder, autism, tic disorders); 29% of children age 6-11 met the exclusion criteria; 25% of adolescents age 12-17 met the exclusion criteria. The application of the exclusion to the measure reduced rates on average across plans by less than 2% for those age 0-5, increased rates by less than 2% for those age 6-11 and increased rates by just over 2% for those age 12-17 (see Table below).

Exclusion for Diagnosis during Measurement Year that has FDA Indication for Antipsychotics

	Age 0-5			Age 6-11			Age 12-17		
	# with exclusion diagnosis	Rate with exclusion	Rate without exclusion	# with exclusion diagnosis	Rate with exclusion	Rate without exclusion	# with exclusion diagnosis	Rate with exclusion	Rate without exclusion
Plan 1	0	25.0%	25.0%	53	63.2%	64.9%	102	73.6%	73.8%
Plan 2	3	75.0%	42.9%	8	76.9%	70.6%	23	77.8%	79.1%
Plan 3	4	33.3%	23.1%	26	35.2%	42.1%	59	54.2%	55.9%
Plan 4	0	NA	NA	4	76.5%	76.2%	13	64.0%	65.8%
Plan 5	0	NA	NA	12	75.0%	66.7%	23	65.8%	69.6%
Plan 6	2	25.0%	50.0%	9	59.4%	58.5%	21	51.2%	55.2%
Plan 7	0	100.0%	100.0%	7	64.3%	71.4%	23	57.4%	68.6%
Plan 8	0	0.0%	0.0%	4	61.5%	70.6%	20	67.8%	72.2%
Plan 9	0	50.0%	50.0%	6	18.2%	35.3%	15	50.0%	52.7%
Plan 10	0	0.0%	0.0%	10	66.7%	60.0%	10	69.8%	71.4%
Plan 11	1	60.0%	66.7%	4	81.8%	66.7%	17	70.7%	70.7%
Plan 12	1	33.3%	25.0%	4	54.5%	66.7%	18	57.1%	60.9%
Plan 13	1	NA	100.0%	5	71.4%	58.3%	10	50.0%	47.5%
Total	**12**	**40.0%**	**38.3%**	**152**	**57.7%**	**59.5%**	**354**	**63.7**	**66.0%**

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*)
2019 Submission: This measure intends to evaluate the use of psychosocial care as a safer first-line therapy for individuals who do not have a primary indication for antipsychotic use. Thus, the exclusion removes individuals with an FDA-indication for antipsychotic use. Evaluation of the exclusion rate across plans indicates consistent and as expected identification of the excluded population across reporting entities. The rates of antipsychotics usage amongst those with an FDA-indication–the excluded population–also match what the literature has indicated, which were found to be approximately 30% (Sohn et al., 2016).

Additionally, experts on our measurement advisory panels recommended specifying the exclusions in the measure based on clinical rational from an accountability perspective, and because it is feasible to collect the data with minimal burden.

Sohn, M., Moga, D.C., Blumenschein, K., & Talbert, J. (2016). National trends in off-label use of atypical antipsychotics in children and adolescents in the United States. Medicine, 95(23), e3784. doi: 10.1097/MD.0000000000003784

2015 Submission: The exclusions in this measure are designed to focus the measure on children in whom psychosocial care is recommended as first-line treatment. The exclusions did not adversely impact the denominator of the measure. Because the exclusions can be collected administratively, they do not pose an undue burden.

_____

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

Plan Type	N	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Commercial	188	55.1%	10.6%	41.0%	48.3%	54.8%	62.6%	69.6%	14.3%	<0.0001
Medicaid	139	57.6%	15.5%	36.4%	52.7%	60.6%	66.6%	75.0%	13.9%	<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 8 MAX States (2008 data)

Mean Rate	10th	25th	50th	75th	90th	IQR
48.2	36.4	37.9	46.9	58.6	61.5	20.7

IQR: Interquartile range

Variation in Performance Rates across 17 Medicaid Plans from one State (2010 data)

Mean Rate	10th	25th	50th	75th	90th	IQR
44.7	27.8	30.4	43.5	54.1	65.4	23.7

IQR: Interquartile range

Variation in Performance Rates across 13 Commercial Plans Nationwide (2012 data)

Mean Rate	10th	25th	50th	75th	90th	IQR
61.8	49.4	53.3	65.8	69.8	71.7	16.5

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. Across both commercial and Medicaid plans, the difference between the 25th and 75th percentile (better performance) is statistically significant.

2015 Submission: The results show that there is a 23.7% gap in performance between Medicaid plans at the 25th and 75th percentiles, a 16.5% gap in performance among commercial plans and a 20.7% gap in performance among states at the 25th and 75th percentiles. This means that states at the 25th percentile have on average 172 less children and adolescents getting recommended first-line psychosocial care than states at the 75th percentile. For Medicaid plans, those at the 25th percentile have on average 119 less children and adolescents getting recommended first-line psychosocial care than plans at the 75th percentile. For commercial plans, those at the 25th percentile have on average 18 less children and adolescents getting recommended first-line psychosocial care 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 Ratings https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ Annual State of Health Care Quality https://www.ncqa.org/hedis/measures/use-of-first-line-psychosocial-care-for-children-and-adolescents-on-anti-psychotics/ Medicaid Child Core Set https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-and-child-health-care-quality-measures/child-core-set/index.html#ChildCoreSet Health Plan Ratings https://www.ncqa.org/hedis/reports-and-research/ratings-2019/ Annual State of Health Care Quality https://www.ncqa.org/hedis/measures/use-of-first-line-psychosocial-care-for-children-and-adolescents-on-anti-psychotics/ Medicaid Child Core Set https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-and-child-health-care-quality-measures/child-core-set/index.html#ChildCoreSet 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/use-of-first-line-psychosocial-care-for-children-and-adolescents-on-anti-psychotics/

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.
Medicaid Child Core Set: This measure is used in the Child Core Set administered by the Centers for Medicare and Medicaid Services (CMS). The Child Core Set intends to evaluate the quality of care and outcomes among children in Medicaid and Children's Health Insurance Program (CHIP). The measure set is reported by over 25 states annually and performance is available publicly at Medicaid.gov.

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 Medicaid Child Core Set, 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 remained largely consistent across Medicaid and commercial health plans (see section 1b.2 for summary of data). This measure requires coordination of care across settings and providers, which requires alignment of resources and infrastructure. While performance rates have remained consistent during the initial years of measure implementation, mechanisms to bridge fragmentation and gaps in care coordination are a national priority and take time to reflect in measurement of specialized populations. In 2019, a total of 188 commercial health plans and 139 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.

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

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

N/A

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, CenCal Health

Harmon Jordan, ScD, Westat

Virginia Raney, LCSW, Center for Medicaid and CHIP Services

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

Laurie Spoll, Aetna

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 Medical Center

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

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

Susan Castellano, Minnesota Department of Human Services

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: 06, 2020

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

Ad.5 When is the next scheduled review/update for this measure? 12, 2016

Ad.6 Copyright statement: The HEDIS® measures and specifications were developed by and are owned by the National Committee for Quality Assurance (NCQA). The HEDIS measures and specifications are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties, or endorsement about the quality of any organization or physician that uses or reports performance measures and NCQA has no liability to anyone who relies on such measures or specifications. NCQA holds a copyright in these materials and can rescind or alter these materials at any time. These materials may not be modified by

anyone other than NCQA. Anyone desiring to use or reproduce the materials without modification for a non-commercial purpose may do so without obtaining any approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA.

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.