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 #: 2799
Measure Title: Use of Multiple Concurrent Antipsychotics in Children and Adolescents
Measure Steward: National Committee on Quality Assurance
Brief Description of Measure: The percentage of children and adolescents 1–17 years of age who were on two or more concurrent antipsychotic medications.
Developer Rationale: This measure addresses inappropriate prescribing patterns 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, less is known about the safety and effectiveness of antipsychotic prescribing patterns in community use (e.g., combinations of medications). Risks of multiple concurrent antipsychotics in comparison to monotherapy have not been systematically investigated. Existing evidence about the harms of multiple concurrent antipsychotic use in children appears largely in case reports and includes increased risk of serious drug interactions, delirium, serious behavioral changes, cardiac arrhythmias and death.

Numerator Statement: Children and adolescents who are on two or more antipsychotic medications concurrently for at least 90 days.
Denominator Statement: Children and adolescents who received 90 days or more of continuous antipsychotic medication treatment.
Denominator Exclusions: N/A

Measure Type: Process
Data Source: Administrative claims
Level of Analysis: Health Plan, Integrated Delivery System, Population : State

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date: N/A

Preliminary Analysis

The preliminary analysis was developed in response to recommendations from NQF's Consensus Task Force and measurement stakeholders as a way to enhance and streamline the measure evaluation and voting processes. The preliminary analysis, developed by NQF staff, will help to guide the Standing Committee evaluation of each measure by summarizing the measure submission and identifying topic areas for discussion. **NQF staff would like to stress that the preliminary analysis is intended to be used as a guide to facilitate the Committee's discussion and evaluation.**

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process* 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.

The developer provides the following evidence for this process measure:

- The rate in this measure (multiple concurrent antipsychotics) relates to the desired outcome (optimal mental and physical outcomes) in the following way: Healthcare provider does not prescribe multiple concurrent antipsychotics >>> Patient receives safer treatment for psychiatric condition present >>> Patient avoids adverse side effects associated with use of multiple concurrent antipsychotic medications >>> Patient experiences improvement in mental and physical outcomes (desired outcome).

- The developer states that "The specific recommendation upon which this measure is based addresses the use of multiple antipsychotics concurrently and notes that the use of multiple antipsychotics has not been studied rigorously and should be avoided. This recommendation is based on established risks of antipsychotics, such as dangerous drug interactions, delirium, serious behavioral changes, cardiac arrhythmias, and death. These risks are in addition to the established side effects of antipsychotic medications that include metabolic disturbance, a serious concern for children."
- The measure is based on clinical practice guidelines. Four guidelines from three organizations are referenced, three of which are ratings. The ratings are:
    - American Academy of Child and Adolescent Psychiatry (AACAP) – not endorsed: ineffective or contraindicated
    - AACAP endorsed best practice principles: Best-practice principles that underlie medication prescribing, to promote the appropriate and safe use of psychotropic medications
    - TMAY Ratings – uses Oxford Centre for Evidence-Based Medicine, guideline is rated C (Level 4 studies or extrapolations from level 2 or 3 studies), very strong recommendation
- While there are several guidelines in this area, the developer focuses on the AACAP guideline since it is most relevant to the measure focus:
    - **Recommendation 8:** "The simultaneous use of multiple concurrent AAAs has not been studied rigorously and generally should be avoided." – Based on a literature review of 147 publications that included clinical trials, meta-analysis, practice guidelines, RCTs, systematic literature reviews, and case reports and series.
    - **Principle 12: "**The prescriber needs a clear rationale for using medication combinations…there is limited evidence in children and adolescents for the use of two antidepressants or two antipsychotics as an initial treatment approach or as a specific endpoint for treatment." – Based on a literature review of 147 publications that included clinical trials, meta-analysis, practice guidelines, RCTs, systematic literature reviews, and case reports and series.
- The developer indicates that the quality of evidence for avoiding multiple concurrent psychotic medications is high.
- No exact estimate exists of the benefits of avoidance of the multiple use of antipsychotic medications, but the short- and long-term risks of these medications in general is well-established.

Questions for the Committee

- *Is the relationship of this measure to patient outcomes supported by the evidence and, if so, how strong is the evidence for this relationship?*
- *The measure specifies concurrent use of medications for 90 days, but the guidelines do not appear to specify a timeframe. Is the timeframe reasonable? Does the Committee wish to explore this further with the developer?*

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

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

The developer provides the following information:

- In a review of the literature, one systematic review found that among youth prescribed any antipsychotic, about one in 10 (9.6 percent, SD 7.2 percent) received multiple concurrent antipsychotics (Toteja et al., 2013). Other studies of multiple concurrent antipsychotics among youth prescribed any antipsychotic have found that prevalence among adolescents is twice that of younger children, and that the rate among adolescents has increased two-fold from the 1990s to the 2000s (Toteja et al., 2013). Another study of a large state Medicaid fee-for-service program found that about 7 percent of children 6–17 years of age on any antipsychotic were prescribed two or more antipsychotics for longer than 60 days (Constantine et al., 2010).
- The developer assessed use of multiple concurrent antipsychotics in Medicaid children using 2008 MAX data from 11 states. It found average rates of 6 percent, with a range of 2.9 to 9.4 percent (a lower rate indicates better care). For children in foster care, the average rate was 6.8 percent, with a range of 1.9 to 10.6 percent. In additional field-testing in Medicaid health plan data from one state, the average percentage of children 0-20

years with use of multiple concurrent antipsychotics was 4.4 percent, with a range of 1.8 percent to 7.0 percent.
- Disparities were noted. In particular, eight states at higher rates of multiple concurrent antipsychotic use in the foster care population compared with the general Medicaid population. Use was higher in adolescents than younger children.
- In both the general and foster care populations, rates were higher for black children and adolescents than Hispanic and white children. For the general population, rates were higher for metropolitan children as opposed to children in rural areas, but for the foster care population, higher rates were seen in rural areas.

Questions for the Committee

- *Is there a gap in care that warrants a national performance measure?*
- *Should this measure be indicated as disparities sensitive? (NQF tags measures as disparities sensitive when performance differs by race/ethnicity [current scope, though new project may expand this definition to include other disparities [e.g., persons with disabilities]).*

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence.

- Supported. The concurrent use for 90 days makes sense.
- While there is evidence to say polypharmacy has increased risk, there is little evidence to say that high dose of 1 medication (4 mg risperdal) is safer than low doses of 2 medications (1 mg risperdal in am and 25 mg seroquel at night)... while this is not the norm, there may be rare clinical times where it is more appropriate and potentially safer and more effective. in general, i imagine these complex cases get referred to psychiatrists.
- "The context of this measure makes it inherently complex. While it's clear there is professional consensus that more than one drug is to be avoided, on a case-by-case basis this is likely to be considered in a small number of children with quite significant mental health conditions.
- The expert opinion and professional consensus appears to be strong given the guidelines referenced, but the direct evidence to support is much weaker. Some specific examples:
- Even the guidelines themselves state that ""use of multiple concurrent AAA's has not been studied rigorously and generally should be avoided."" This is not direct evidence of harm, but rather absence of evidence of safety. Their finding is that the use (of multiple AAA's is ""not endorsed,"" which speaks to lack of evidence rather than affirmative evidence of harm.
- Principle 12 only says that ""prescriber needs a clear rationale for using medication combinations…"" and that the principle applies to prescribing ""as an initial treatment approach."" I don't think the measure limits itself to initial treatment.
- The developers refer to a review of 147 studies, but it is unclear how many of these really relate to the specific point of harm from >1 antipsychotic. The AACAP-AAA review did not produce estimates of the benefit of avoidance. In sec 1.c.3 the developer state: ""Risks of multiple concurrent antipsychotics in comparison to mono therapy have not been systematically investigated."" In summary, I do not agree with the statement from the developers that the ""quality of the evidence in support of avoiding multiple concurrent antipsychotic medications is high.""""
- Measure of a process - I don't see clear evidence described that links concomitant use directly to poor outcomes for the child. There are myriad studies describing adverse effects of antipsychotic use in children, however, so it reasons that concomitant use will amplify these effects. Several clinical practice guidelines cited which directly address the focus of this measure. I suspect 90 days was selected to allow for some 'washout' period if a child is being transitioned from one antipsychotic to another. While individual practice will vary on the timeline to transition from one medicine to another, 90 days is sufficient time for any transition to have occurred.
- While there is little research on the use of multiple antipsychotics with children and adolescents, it is well established that the use of antipsychotics can increase metabolic disorder, cardiac issues, behavioral changes, and other significant problems.
- The Measure proposed looks to support the contention that multiple concurrent antipsychotic use in Children and Adolescents may result in numerous negative affects. This measure and the collection of data proposed if this Measure is endorsed has a goal to decrease the use of multiple concurrent antipsychotic drugs in these populations. the Developers state strong evidence against this multiple drug use. The evidence relates well to the process

Measure proposed and is supported by the stated rationale. Recommendation 8 describes the evidence used to support this measure.

1b. Performance Gap.

- There is overall less than optimal performance. The developer stated there were racial disparities, but did not provide specific numbers.
- Performance gap exists. the measure describes disparities by population subgroups (race), but may also be explained by other psychosocial risk factors (high stress home environment, lack of supports/resources, attachment issues, etc.).
- A performance gap is demonstrated by variability in rates seen at the state level (using Medicaid MAX data, 2008)-- they were overall 6% with a range of 2.9 to 9.4. Similar ranges were seen at the health plan level (Medicaid plans from one state). The mean rate is not "0" but given the concerns above, it's not clear what the right number is that would balance risks and benefits. Especially when it's possible that some children have already failed single medication treatment. While the relative difference are great (3-fold between lowest and highest) the absolute differences are less dramatic. However, there does appear to be implicit consensus of experts that the current rate is too high.
- Disparities are demonstrated by race/ethnicity and by age (African American children and adolescents more likely to receive >1 agent).
- Performance data is provided, including a systematic review, a single state review of Medicaid FFS data and an 11-state MAX data review. There is a clear gap in care and data cited to suggest this gap is worsening over time. There appear to be clear disparities in certain subpopulations (black children, adolescents, children in foster care) and I would indicate this as disparities sensitive.
- The rate of using multiple antipsychotics with this age group has doubled in the past twenty years.
- There appears to be higher utilization of multiple antipsychotics in minority youth and those in foster care.
- Performance data on the measure was provided including state, health plans and other data. Data at the state level is cited. For example, the Developer looked at Medicaid recipients, comparing those in Foster Care and total Medicaid recipients. in these cases higher use of multiple drugs in Foster Care in 8 states. Also, higher use in Black patients than other groups is seen in some groups. From their discussion a gap in care is seen comparing Medicaid to other groups.

Criteria 2: Scientific Acceptability of Measure Properties
2a. Reliability
2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- The numerator for this measure is children and adolescents who are on two or more antipsychotic medications concurrently for at least 90 days. The denominator is children and adolescents who received 90 days or more of continuous antipsychotic medication treatment.
- The numerator and denominator details provide steps to identify patients for inclusion and include a list of medications. No codes are needed to calculate the measure.
- The measure is stratified by age, but is not risk adjusted.

Questions for the Committee.

- *Are all appropriate medications included?*
- *Is the logic or calculation algorithm clear?*

- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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.

- This measure was tested at the performance measures score level using a beta-binomial signal-to-noise analysis. For this type of testing, a score of zero implies that all the variability in a measure is attributable to measurement error. A score of 1.0 implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one reporting entity from another. A score of 0.7 or higher indicates adequate reliability to distinguish performance between two entities and is considered acceptable.
- Per the NQF algorithm, reliability testing at the computed performance measure score may be rated HIGH, MODERATE, or LOW depending on the testing results.
- The developer reports the following testing results:
  - The average state level reliability was 0.99, and the minimum was 0.96, suggesting high reliability at the state level.
  - The reliability for Medicaid health plans averaged 0.64, with a minimum of 0.28.
  - The reliability for commercial health plans averaged 0.42 average, with a minimum of 0.08.

Questions for the Committee

- *The developer concludes the measure is reliable only at the state level. Does the Committee concur?*

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

- The specifications are consistent with the evidence. The goal of the measure is to assess inappropriate prescribing of antipsychotic medication to children and adolescents. The evidence provided supports the specifications.

Question for the Committee

- *Are the appropriate medications included in the specifications?*

2b2. Validity testing

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.

- The measure was tested at the performance measure score level using both empirical testing and face validity.
- For the empirical testing, the developer assessed construct validity with two types of analyses: correlations among measures and rankings of health plans and states on measures on the three antipsychotic medication measures.
  - Correlations were tested using health plans, as there was not enough entities to test between states. The results showed that among Medicaid health plans, there were no statistically significant correlations between the *Multiple Concurrent* measure and the other measures addressing antipsychotic use in children and adolescents. Among national commercial plans, there was moderate negative correlation between the *Follow-up Visit* and *Multiple Concurrent* measures (r=-.58, p=0.02).
  - The developer states that "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." Their interpretation is that the results show that plans and states can be approximately ranked based on profiles of performance across multiple measures. The consistent performance across measures suggest the measures are assessing a dimension of quality.
- Per the NQF algorithm, validity testing at the computed performance measure score may be rated HIGH,

MODERATE, or LOW depending on the testing results.

- The developer used its standardized HEDIS process to test face validity of the measure construct, but does not explicitly call out face validity of the computed performance score, as required by NQF.
  - The developer worked with five expert panels to identify the most appropriate method for assessing the use of multiple concurrent antipsychotics among this patient population. All of the panels concluded this measure was specified to assess multiple concurrent use of antipsychotics.
  - The draft measure was put out for public comment and brought to the developer's Committee on Performance Measurement.
  - The developer states that the measure has sufficient face validity.

Questions for the Committee

- *Do the results of the empiric testing demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you believe that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- There are no exclusions.

Questions for the Committee

- *Should there be any exclusions for this measure?*
- *Does the Committee believe there are other threats to validity?*

2b4. Risk adjustment:

- The measure is not risk adjusted.

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer states that the results indicate that there is 2.1% gap in performance between Medicaid plans at the 25th and 75th percentiles, a 3.2% gap in performance among commercial plans and a 4.4% gap in performance among states at the 25th and 75th percentiles. This means states at the 75th percentile have on average 504 more children and adolescents receiving multiple concurrent antipsychotics than states at the 25th percentile.

Question for the Committee

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- This is not needed.

2b7. Missing Data

- The measure is collected using all administrative data sources. According to the developer there are no missing data, so this is not applicable.

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. Specifications

- Would like clarification on the 90 day continuous use specification for numerator hits. Must the child be on the same two for 90 consecutive days for both, or would a child who is on one for 90 days and a second one for the first 45 days and a different "second" one for the remaining 45 days be included in the numerator.
- I'm curious if there is one scheduled med (ability for example), and one prn med (risperdal m-tab as needed for agitation) - if that gets included in numerator.
- The language of the numerator and denominator could be more precise. Children who are "on" these medications is imprecise. Rather, the numerator (and denominator) are based on pharmacy dispensings covering a 90-day period. The numerator calculation is complex but understandable. I wonder if there is a typo: it says that if the number of days between the end date (of dispensing 1) and the start day of dispensing 2 "= 15" days, the gap days should be counted. My guess is that this should say "<=15 days".

It is not clear how the generic drug names are to be translated into the measure calculation. Should there be a list of NDC codes, or is it left to each plan to determine how to capture these.
For the denominator, it's not clear why a 32 day gap is allowed (compared to a 15 day gap in the numerator).
Review for whether all appropriate medications are included must include content experts (e.g. pharmacists), or reliance on the process used by the developers.

- Numerator and denominator are clearly defined. Appropriate medications are included. The calculation algorithm is clear and appears it can be consistently implemented.
- The measure specifies both first and second generation antipsychotics and appears thorough.
- Reliability specifications were submitted in detail. It is my understanding that codes with descriptors were not provided. One presented analysis in which a score of 1.0 indicates high reliability was presented. This Measure was shown to have a score of .99 average at the state level (.96 minimum) indicating high reliability for the use of this Measure. It appears to me that this measure can be consistently implemented.

2a2. Reliability testing

- Reliability was tested with the MAX data set (11 states), 17 Medicaid health plans within one state, and a sample of commercial plans. The method for reliability testing is the beta-binomial-signal to noise method, which is appropriate. The reliability is acceptable in the very large state-level analysis. But, it is of borderline acceptability in the Medicaid plans (that have larger sample sizes) but was not acceptable (minimum reliability .08) in some of the commercial plans. This is directly related to the number of children meeting denominator criteria. For example, 24 of 72 commercial plans had less than 30 children (so were excluded). 25% of the included plans had less than 90. This highlights the need for using this measure only in settings with sufficient samples of children meeting the denominator criteria. By the algorithm I would rate the reliability as Low at the health plan level and Moderate at the state level.
- Calculations suggest reliability is high only at the state level.
- My understanding of this work is limited. It does appear that reliability was completed and it was determined that measurements were reliable at the state level only.

2b1. Validity Specifications

- What does the developer attribute to the disparity between the reliability for Medicaid/commercial health plans compared to state level.
- The measure specifications are consistent with the intent, and consistent with the evidence, at the level presented-- with the caveats above. The list of medications appears reasonable, but requires review by individuals with content expertise.
- Appropriate medications are included.
- The testing suggests that this measures is a valid measure for assessing the rate at which providers prescribe more than one antipsychotic medication in youth.
- The reliability testing evidenced a score of .9, suggesting the measure is highly reliable at detecting differences at the state level. Reliability estimates for health plans was significantly lower.
- It does not appear to me that there are specifications inconsistent with the evidence. It also appears that the target population (children and adolescents) values and would be served by more consistent avoidance of multiple antipsychotic drug use if this measure is endorsed.

2b2. Validity Testing

- There is good face validity on the measure based on the developers use of five expert panels and opportunity for public comment. The empirical validity studies are less convincing. Correlation with other measures in this topic area is is poor using Spearman correlation coefficients. The ranking method suggests only rough stability of rankings (of health plans and states) across related measures. In a sense, this is as much about reliability as validity, and is difficult to interpret with no quantification of what would be considered "good" validity. I would consider the results of the empirical testing Low, but the face

validity as Moderate.

- Developers demonstrate consistency with this measure in comparison to other antipsychotic measures, as well as standardized method to demonstrate face validity. Score on this measure is an indicator of quality.
- The measure appears to distinguish between low and high performer states well.
- Measure was ranked as a high priority by expert panel (face validity).
- As with 2a2, my understanding of this work is limited. Validity was tested on several levels.

2b3-2b7. Threats to Validity

- Because this is based on pharmacy claims, it's likely the data are complete. However, some caveats need to be made. It relies on patients having pharmacy benefits, and always using the same plan for dispensings. Because the numerator and denominator both rely on this, I do not believe it is a major concern. There is no information on medicines that the patient did not take (or were discontented by the prescriber). The developers should address whether any of these are concerns.
- Developers state this is administrative data, and therefore there are no missing data.
- Authors indicate no missing data
- There were no exclusions included in this measure. 2b5 - Gaps were seen that this measure proposed to decrease.

Criterion 3. Feasibility

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

The developer notes:

- These elements are all generated through normal process of care and are in defined fields in electronic claims.
- The measure is a part of HEDIS, which has a standardized collection and calculation process, as well as a system to collect real-time feedback from measure users.
- Field testing results showed the measure is feasible to be collected by health plans and states using administrative claims data.
- As part of HEDIS, the data elements are subject to that program's data collection and audit requirements.
- This is not an eMeasure.

Questions for the Committee

- *Are the required data elements routinely generated and used during care delivery?*
- *Does the testing data collection strategy indicate the measure is ready to be put into operational use?*

Committee pre-evaluation comments

Criteria 3: Feasibility

- While polypharmacy is not desired, there may be relatively rare clinical cases where it is justified and does not represent poor quality of care. in general, complex cases should involve specialists.
- Feasibility of the measure is good, given that it is based in pharmacy claims only, and is currently being used (on a voluntary basis) as a HEDIS measure. Usability is good based on these as well.
- All required data elements are routinely generated and used in the course of normal care delivery. No concerns about putting this measure into operational use.
- Feasible because it relies on administrative claims data. Measure can be obtained through data that is secured through routine daily care.
- The data elements required in this measure are routinely generated and used during care delivery. No

concerns regarding collection strategy.

Criterion 4: Usability and Use

4. Usability and 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.

- The measure is currently in use in for both quality improvement with benchmarking and public reporting.
- It is included in Quality Compass for Medicaid 2015, a tool that displays health plan-level performance rates for HEDIS measures. It is used for selecting a health plan, conducting competitor analysis, examining quality improvement and benchmarking plan performance.
- The measure also is reported on in The State of Health Care Quality Report, a national report produced by the developer including the results from HEDIS measures.
- This is a new measure and improvement results are not yet available.
- No unintended consequences have been reported thus far.

Question for the Committee

- *Do the benefits of the measure outweigh any potential unintended consequences?*

Committee pre-evaluation comments

Criteria 4: Usability and Use

- Would like to hear developers provide perspective on HEDIS 2015 analysis for this measure. Are the rates cited in the results of the testing of the measure rate per 1,000 or rate per 100.
- The measure has been approved for use in the Quality Compass for Medicaid.

Criterion 5: Related and Competing Measures

- This measure, 2799, is related to one NQF-endorsed measure, 2337: Antipsychotic Use in Children Under 5 Years Old.
- This measure has a different target population of those who have continuous use of antipsychotics for 90 days or more, includes more children (up to age 18 years), and has a different focus (i.e., a specific type of non-recommended practice [multiple concurrent use] as opposed to any use).

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** N/A

Measure Title: Use of Multiple Concurrent Antipsychotics in Children and Adolescents

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

Date of Submission: 10/9/2015

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

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

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

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

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

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

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

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

Outcome

☐ Health 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

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

☒ Process: Multiple concurrent antipsychotic medication avoided for those with continuous antipsychotic medication treatment

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

N/A

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

N/A

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

The rate in this measure (multiple concurrent antipsychotics) relates to the desired outcome (optimal mental and physical outcomes) in the following way:

Health care provider does not prescribe multiple concurrent antipsychotics >>> Patient receives safer treatment for psychiatric condition present >>> Patient avoids adverse side effects associated with use of multiple concurrent antipsychotic medications >>> Patient experiences improvement in mental and physical outcomes (desired outcome).

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

- American Academy of Child and Adolescent Psychiatry. Practice Parameter for the Use of Atypical Antipsychotic Medications in Children and Adolescents. http://www.aacap.org/App_Themes/AACAP/docs/practice_parameters/Atypical_Antipsychotic_Medications_Web.pdf (July 12, 2012)
- American Academy of Child and Adolescent Psychiatry. September 2009. Practice parameter on the use of psychotropic medication in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry.* 48(9):961–73.
- Scotto, Rosato N., C.U. Correll, E. Pappadopulos, A. Chait, S. Crystal, P.S. Jensen. June 2012. Treatment of maladaptive aggression in youth: CERT guidelines II. Treatments and ongoing management. *Pediatrics.* 129(6):e1577–86.
- Texas Department of Family and Protective Services and University of Texas at Austin College of Pharmacy. 2013. Psychotropic Medication Utilization Parameters for Foster Children. http://www.dfps.state.tx.us/documents/Child_Protection/pdf/TxFosterCareParameters-September2013.pdf (October 22, 2013)

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Guideline (Date)	Population	Recommendation or Statement	Type/Grade
AACAP-AAA (2011) Practice parameter for the use of atypical antipsychotic medications in children and adolescents	5-18 years	"The simultaneous use of multiple concurrent AAAs has not been studied rigorously and generally should be avoided." (Recommendation 8)	Not Endorsed
AACAP-PsyMed (2009) Practice parameter on the use of psychotropic medication in children and adolescents	≤18 years	"The prescriber needs a clear rationale for using medication combinations….there is limited evidence in children and adolescents for the use of two antidepressants or two antipsychotics as an initial	Best practice principle

		treatment approach or as a specific endpoint for treatment." (Principle 12)	
TMAY (2012) Center for Education and Research on Mental Health Therapeutics—Treatment of maladaptive aggression in youth	≤18 years	Use of two simultaneous psychotropic medications should be avoided (Recommendation 18)	Evidence: C Strength of Recommendation: Very Strong
TX (2010) Texas Department of Family and Protective Services – Psychotropic medication utilization parameters for foster children	Children (age un-specified)	Prescribing multiple antipsychotics is a situation that warrants clinical review.	Not specified*

*TX (2010) did not specify the use of a rating system.

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

Guideline Developer	Definition
AACAP	**Not endorsed:** Ineffective or contraindicated.
AACAP endorsed best-practice principles	**Best-practice principles** that underlie medication prescribing, to promote the appropriate and safe use of psychotropic medications
TMAY Ratings	Oxford Centre for Evidence-Based Medicine grade of evidence (A-D) C: Level 4 studies or extrapolations from level 2 or 3 studies
	Strength of Recommendation: Very strong (≥90% agreement)

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Guideline Developer	Definition
AACAP	**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
	Clinical guidelines: Strong empirical evidence (non-randomized controlled trials, cohort or case-control studies), and/or strong clinical consensus; expected to apply in most cases (75% of the time)
	Options: Acceptable but not required; there may be insufficient evidence to support higher recommendation (uncontrolled trials, case/series reports)

Guideline Developer	Definition
TMAY Ratings	Oxford Centre for Evidence-Based Medicine grade of evidence (A-D)
	A: Consistent level 1 studies
	B: Consistent level 2 or 3 studies or extrapolations from level 1 studies
	D: Level 5 evidence or troublingly inconsistent or inconclusive studies of any level
	Strength of Recommendation: Strong (70-89% agreement)
	Strength of Recommendation: Fair (50-69% agreement)
	Strength of Recommendation: Weak (<50% agreement)

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

OCEBM Levels of Evidence Working Group. 2011. The Oxford 2011 levels of evidence. http://www.cebm.net/index.aspx?o=5653 (October 12, 2013)

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Several guidelines address the use of multiple concurrent antipsychotics in children and adolescents; each guideline cautions against the use of multiple concurrent antipsychotics given the lack of evidence supporting this type of use. While we list the full range of guidelines in sections 1a.4.2 and 1a.4.3 above, we focus on and describe in more detail the American Academy of Child and Adolescent Psychiatry (AACAP) Guideline in the remaining sections, as it is most closely relevant to the specified measure. The AACAP guideline addresses the use of antipsychotic medications in children and adolescents. The specific recommendation upon which this measure is based addresses the use of multiple antipsychotics concurrently and notes that the use of multiple antipsychotics has not been studied rigorously and should be avoided. This recommendation is based on established risks of antipsychotics, such as dangerous drug interactions, delirium, serious behavioral changes, cardiac arrhythmias, and death. These risks are in addition to the established side effects of antipsychotic medications that include metabolic disturbance, a serious concern for children.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

See table under 1a.4.2 for the level of evidence grade given to each guideline. See table under 1a.4.3 for the definition of the level of evidence grade given to each guideline.

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

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

See table under 1a.4.4 for the definition of the level of evidence grade not given to the guidelines.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 1990-2010

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

The guidelines listed in our table above address antipsychotic polypharmacy among children. The AACAP-AAA recommendation is rated a "Not Endorsed," indicating there is no rigorous/substantial empirical evidence and/or overwhelming clinical consensus to support prescribing multiple concurrent antipsychotics for children and adolescents. The guideline states that "due to the possibility of significant risks associated with these agents [atypical antipsychotics], the use of more than one agent is not recommended and is not supported in the scientific literature". AACAP includes an additional, broader guideline around use of multiple concurrent psychotropic medications in youth; we focus on the antipsychotic-specific AACAP-AAA guideline here and describe the body of evidence for each relevant recommendation below.

When developing their guidelines, AACAP limited its evidence review to clinical trials, meta-analysis, practice guidelines, randomized controlled trials (RCTs), systematic literature reviews, and case reports and series. AACAP selected a total of 147 publications for careful examination based on their weight in the hierarchy of evidence attending to the quality of individual studies, relevance to clinical practice and the strength of the entire body of evidence. 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.

Recommendation 8: "The simultaneous use of multiple concurrent AAAs has not been studied rigorously and generally should be avoided."

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.

Principle 12: "The prescriber needs a clear rationale for using medication combinations….there is limited evidence in children and adolescents for the use of two antidepressants or two antipsychotics as an initial treatment approach or as a specific endpoint for treatment."

This principle 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.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

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 in support of avoiding multiple concurrent antipsychotic medication is high.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

The AACAP-AAA review did not include an exact estimate of benefits of avoiding multiple concurrent antipsychotics in youth. However, the evidence has established that use of multiple concurrent antipsychotic is associated with adverse short-term psychotic, behavioral, cardiovascular, and other side effects in youth and to negative long-term health outcomes throughout the lifespan.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

The AACAP review did not examine the potential harms of avoiding multiple concurrent antipsychotics in youth, which have not been thoroughly investigated. However, the harms of unnecessary antipsychotic use in general 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; Safer et al., 2003; Srinivasan et al. 2002; Van Bennekom et al., 2013).

<u>Citations</u>

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.

Safer, D.J., J.M. Zito, S. DosReis. 2003. Concomitant psychotropic medication for youths. *American Journal of Psychiatry.* 160(3): p. 438–49.

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.

Van Bennekom, M., H. Gijsman, F. Zitman. 2013. Antipsychotic polypharmacy in psychotic disorders: A critical review of neurobiology, efficacy, tolerability and cost effectiveness. *Journal of Psychopharmacology.* 27: 327.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

To our knowledge, there have been no new studies that contradict the current body of evidence.

1a.8 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.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

1. Evidence, Performance Gap, Priority – 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 subcriteria to pass this criterion and be evaluated against the remaining criteria.***

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

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., the benefits or improvements in quality envisioned by use of this measure)*
This measure addresses inappropriate prescribing patterns 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, less is known about the safety and effectiveness of antipsychotic prescribing patterns in community use (e.g., combinations of medications). Risks of multiple concurrent antipsychotics in comparison to monotherapy have not been systematically investigated. Existing evidence about the harms of multiple concurrent antipsychotic use in children appears largely in case reports and includes increased risk of serious drug interactions, delirium, serious behavioral changes, cardiac arrhythmias and death.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. 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 included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
New measure: not applicable

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.**
In a review of the literature, one systematic review found that among youth prescribed any antipsychotic, about one in 10 (9.6 percent, SD 7.2 percent) received multiple concurrent antipsychotics (Toteja et al., 2013). Other studies of multiple concurrent antipsychotics among youth prescribed any antipsychotic have found that prevalence among adolescents is twice that of younger children, and that the rate among adolescents has increased two-fold from the 1990s to the 2000s (Toteja et al., 2013). Another study of a large state Medicaid fee-for-service program found that about 7 percent of children 6–17 years of age on any antipsychotic were prescribed two or more antipsychotics for longer than 60 days (Constantine et al., 2010).

As part of the measure's field-testing, we assessed use of multiple concurrent use of antipsychotic medications in Medicaid children, using the Medicaid Analytic eXtract (MAX) data files. Analysis of administrative claims data from 11 states demonstrated that the average percentage of children with use of multiple concurrent antipsychotics was 6.0 percent, with a range of 2.9 to 9.4 percent (a lower rate indicates better care). For children in foster care, the average rate was 6.8 percent, with a range of 1.9 to 10.6 percent. In additional field-testing in Medicaid health plan data from one state, the average percentage of children 0-20 years with use of multiple concurrent antipsychotics was 4.4 percent, with a range of 1.8 percent to 7.0 percent.

Citations

Constantine, R., M. Bengtson, T. Murphy, et al. 2012. Impact of the Florida Medicaid Prior-Authorization Program on use of antipsychotics by children under age six. Psychiatric Services. 12: DOI: 10.1176/appi.ps.201100346.

Toteja, N., J.A. Gallego, E. Saito, et al. 2013. Prevalence and correlates of antipsychotic polypharmacy in children and adolescents receiving antipsychotic treatment. International Journal of Neuropsychopharmacology. DOI: 10.1017/S1461145712001320.

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 endorsement maintenance. 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 subcriterion on improvement (4b.1) under Usability and Use.*
We tested the measure and evaluated disparities in its performance using Medicaid Analytic eXtract (MAX) data. We assessed performance by age group (0-5, 6-11, 12-20), race/ethnicity, foster care status, and rurality/urbanicity.

MAX DATA DESCRIPTION AND RESULTS
Our MAX dataset was composed of 2008 service data from 11 states. The analysis population included all Medicaid enrollees aged 0-20 on December 31, 2008 in the 11 states. Both fee-for-service and managed care enrollees were included. Data files included person summary, outpatient claims, inpatient claims and prescription claims. States were chosen due to completeness of their data for managed care enrolled beneficiaries.

Of the 11 states, eight had higher rates of multiple concurrent antipsychotic use in the foster care population compared with the general Medicaid population. In both the general population and foster care population, multiple concurrent antipsychotic use was highest among adolescents compared with the lower age strata. In the general population, rates of multiple concurrent antipsychotic use were slightly higher among Black Non-Hispanic children and adolescents (7.5 percent) than Hispanic (6.1 percent) and White Non-Hispanic (6.5 percent) children and adolescents. Similarly, in the foster care population, rates of multiple concurrent antipsychotic use were slightly higher among Black Non-Hispanic children and adolescents (8.6 percent) than Hispanic (6.7 percent) and White Non-Hispanic (7.6 percent) children and adolescents. For the general population of children, higher rates of multiple concurrent antipsychotic use were seen in metropolitan areas (6.8 percent) than rural areas (5.7 percent). However, within the foster care population, higher rates were seen in rural areas (9.5 percent), compared with metropolitan areas (6.6 percent).

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
A leading cause of morbidity/mortality, Patient/societal consequences of poor quality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Antipsychotic prescribing for children has increased rapidly in recent decades, driven by new prescriptions and by longer duration of use (Patten et al., 2012; Cooper et al., 2006). Although some evidence supports the efficacy of antipsychotics in youth for certain narrowly defined conditions, less is known about the safety and effectiveness of antipsychotic prescribing patterns in community use (e.g., combinations of medications, off-label prescribing, dosing outside of recommended ranges).

Children and adolescents prescribed antipsychotics are more at risk for serious health concerns, including weight gain, extrapyramidal side effects, hyperprolactinemia and some metabolic effects (Correll et al., 2011). Girls treated with certain antipsychotics may also be at increased risk for gynecological problems (Talib et al., 2013) and osteoporosis (Cohen et al., 2012).

Risks of multiple concurrent antipsychotics in comparison to monotherapy have not been systematically investigated; existing evidence appears largely in case reports, and includes increased risk of serious drug interactions, delirium, serious behavioral changes, cardiac arrhythmias and death (Safer et al., 2003). Research demonstrating that the pharmacokinetics of antipsychotics may vary by developmental stage (Correll et al., 2011) suggests that use of multiple concurrent antipsychotics may pose greater risks for children and adolescents compared to adults.

The financial impact of multiple concurrent antipsychotic use in children has not been examined; however, antipsychotics are a costly

form of drug therapy. Atypical antipsychotics have the greatest mean prescription cost ($132) of any psychotropic medication (Martin & Leslie, 2003) and until recently were the most costly drug class within the Medicaid program (Crystal et al., 2009). Additionally, there are substantial long-term costs of treating side effects associated with antipsychotic medications, including treatment of obesity, diabetes and dyslipidemias. There is some evidence that these health conditions, such as new onset diabetes, do not always resolve after discontinuation of the antipsychotic (Lean and Pajonk, 2003). Although this is an understudied area, it is reasonable to assume that unresolved side effects from antipsychotics would be associated with the long-term increases in health care costs that have been established for obesity and diabetes.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Cohen, D., O. Bonnot, N. Bodeau, et al. 2012. Adverse effects of second-generation antipsychotics in children and adolescents. Journal of Clinical Psychopharmacology. 32:309–16.

Cooper, W.O., P.G. Arbogast, H. Ding, G.B. Hickson, D.C. Fuchs, and W.A. Ray. 2006. Trends in prescribing of antipsychotic medications for US children. Ambulatory Pediatrics. 6(2):79–83.

Correll, C.U., C.J. Kratochvil, J.S. March. 2011. Developments in pediatric psychopharmacology: Focus on stimulants, antidepressants, and antipsychotics. Journal of Clinical Psychiatry. 72:655–70.

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.

Lean, M.E., F.G. Pajonk. 2003. Patients on Atypical Antipsychotic Drugs Another high-risk group for type 2 diabetes. Diabetes Care. 26(5), 1597–605.

Martin, A., D. Leslie. 2003. Trends in psychotropic medication costs for children and adolescents, 1997-2000. Archives of Pediatric Adolescent Medicine. 157(10):997–1004.

Patten, S.B., W. Waheed, L. Bresee. 2012. A review of pharmacoepidemiologic studies of antipsychotic use in children and adolescents. Canadian Journal of Psychiatry. 57:717–21.

Safer, D.J., J.M. Zito, S. DosReis. 2003. Concomitant psychotropic medication for youths. American Journal of Psychiatry. 160(3): p. 438–49.

Talib, H.J., E.M. Alderman. 2013. Gynecologic and reproductive health concerns of adolescents using selected psychotropic medications. Pediatric Adolescent Gynecology. 26:7–15.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

N/A

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 subcriteria 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, Mental Health

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

Safety, Safety : Medication Safety

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.)*
None

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)*
No data dictionary **Attachment:**

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Children and adolescents who are on two or more antipsychotic medications concurrently for at least 90 days.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
12 months (January 1 – December 31)

S.6. **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, 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.
Children and adolescents who are on two or more antipsychotic medications (Table APC-A) concurrently for at least 90 consecutive days during the measurement year (January 1 – December 31).

To identify the numerator: for each patient in the eligible population, by drug, identify all prescription events, start dates and end dates. Then identify the number of concurrent antipsychotic medication treatment events.
Step 1: For each patient, identify the first day during the measurement year where the patient was being treated with two or more different antipsychotic medications; this is the concurrent antipsychotic medication treatment event start date.
Step 2: Beginning with (and including) the start date, identify the number of consecutive days where the patient remains on two or more different antipsychotic medications. If the number of days =90 days, the patient is numerator compliant.
Step 3: If the number of consecutive days on multiple antipsychotic medications is <90 days, identify the end date and identify the next day during the measurement year where the patient was being treated with two or more different antipsychotic medications. If the number of days between the end date and the next start date is =15 days, include the days in the concurrent antipsychotic medication treatment event (concurrent antipsychotic medication treatment events allow for a 15-day gap).
Step 4: If the number of days between the end date and the next start date exceeds 15 days, end the event; using the new start date, continue to assess for concurrent antipsychotic medication treatment events.
Step 5: Continue this process until the number of concurrent antipsychotic medication treatment days is =90 consecutive days (i.e., the patient is numerator compliant) or until the measurement year is exhausted (i.e., no concurrent antipsychotic medication treatment events were identified during the measurement year).

Table APC-A: Antipsychotic Medications
First-generation antipsychotic medications: Chlorpromazine HCL; Fluphenazine HCL; Fluphenazine decanoate; Fluphenazine enanthate; Haloperidol; Haloperidol decanoate; Haloperidol lactate; Loxapine HCL; Loxapine succinate; Molindone HCL; Perphenazine; Pimozide; Promazine HCL; Thioridazine HCL; Thiothixene; Thiothixene HCL; Trifluoperazine HCL; Triflupromazine HCL

Second-generation antipsychotic medications: Aripiprazole; Asenapine; Clozapine; Iloperidone; Lurasidone; Olanzapine; Olanzapine pamoate; Paliperidone; Paliperidone palmitate; Quetiapine fumarate; Risperidone; Risperidone microspheres; Ziprasidone HCL; Ziprasidone mesylate

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Children and adolescents who received 90 days or more of continuous antipsychotic medication treatment.

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Children's Health, Populations at Risk

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, 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)*
Children and adolescents age 1-17 as of December 31 of the measurement year (January 1 – December 31) who received 90 days or more of continuous antipsychotic medication (Table APC-A) treatment.

The instructions outlined here are designed to identify through pharmacy claims children with at least 90 days of continuous antipsychotic use. The measure allows for a 32-day gap in order to account for missed prescription fills, which can be common among children, particularly for off-label use for behavioral control.

Step 1: Identify patients in the specified age range who were dispensed an antipsychotic medication during the measurement year (January 1 – December 31).
Step 2: For each patient, identify all antipsychotic prescriptions during the measurement year. For each drug, identify start and end dates of the prescriptions. Starting with the first prescription in the measurement year determine if there is a second dispense date of that same drug. If there is no second dispensing event with the same Drug ID, the start date is the first prescription's dispense date and the end date is the start date plus the days supply minus one. If there is a second dispensing event of the same drug, determine if there are gap days (a 32-day gap is allowed). Calculate the number of days between (but not including) the first prescription's dispense date and the second prescription's dispense date. If the number of days is less than or equal to the first prescription's days supply plus 32 days, the gap is less than or equal to 32 days and is allowed. The start date is the first prescription's dispense date and the end date is the second prescription's dispense date plus days supply minus one.
Step 3a: Continue assessing all subsequent dispensing events with allowable gaps for the same drug and adjust end dates as needed. If there is a second dispensing event of the same drug and there is a gap that exceeds the allowable gap, assign an end date for this drug event and begin with the next prescription to again assess if there is 90 days of continuous use. A patient can have multiple start and end dates per drug during the measurement year.
Step 3b: Continue assessing each dispensed prescription for each drug until all dispensing events are exhausted. If a dispensing event goes beyond December 31 of the measurement year, assign the end date as December 31.
Step 4: For each patient, identify if they were dispensed at least 90 consecutive treatment days of antipsychotics during the measurement year.

Table APC-A: Antipsychotic Medications
First-generation antipsychotic medications: Chlorpromazine HCL; Fluphenazine HCL; Fluphenazine decanoate; Fluphenazine enanthate; Haloperidol; Haloperidol decanoate; Haloperidol lactate; Loxapine HCL; Loxapine succinate; Molindone HCL; Perphenazine; Pimozide; Promazine HCL; Thioridazine HCL; Thiothixene; Thiothixene HCL; Trifluoperazine HCL; Triflupromazine HCL
Second-generation antipsychotic medications: Aripiprazole; Asenapine; Clozapine; Iloperidone; Lurasidone; Olanzapine; Olanzapine pamoate; Paliperidone; Paliperidone palmitate; Quetiapine fumarate; Risperidone; Risperidone microspheres; Ziprasidone HCL; Ziprasidone mesylate

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
N/A

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, 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)*
N/A

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, 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 with at S.2b)
Report three age stratifications and a total rate:
1–5 years
6–11 years
12–17 years
Total (sum of the age stratifications)

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
N/A

S.16. Type of score:
Rate/proportion
If other:

S.17. 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 = Lower score

S.18. Calculation Algorithm/Measure Logic *(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; aggregating data; risk adjustment; etc.)*
Step 1: Determine the eligible population, or the denominator.
Step 1a: Identify patients in the specified age range who were dispensed an antipsychotic medication during the measurement year (January 1 – December 31).
Step 1b: For each patient, identify all antipsychotic prescriptions during the measurement year. For each drug, identify start and end dates of the prescriptions. Starting with the first prescription in the measurement year determine if there is a second dispense date of that same drug. If there is no second dispensing event with the same Drug ID, the start date is the first prescription's dispense date and the end date is the start date plus the days supply minus one. If there is a second dispensing event of the same drug, determine if there are gap days (a 32-day gap is allowed). Calculate the number of days between (but not including) the first prescription's dispense date and the second prescription's dispense date. If the number of days is less than or equal to the first prescription's days supply plus 32 days, the gap is less than or equal to 32 days and is allowed. The start date is the first prescription's dispense date and the end date is the second prescription's dispense date plus days supply minus one.
Step 1c: Continue assessing all subsequent dispensing events with allowable gaps for the same drug and adjust end dates as needed. If there is a second dispensing event of the same drug and there is a gap that exceeds the allowable gap, assign an end date for this drug event and begin with the next prescription to again assess if there is 90 days of continuous use. A patient can have multiple start and end dates per drug during the measurement year.
Step 1d: Continue assessing each dispensed prescription for each drug until all dispensing events are exhausted. If a dispensing event goes beyond December 31 of the measurement year, assign the end date as December 31.
Step 1e: For each patient, identify if they were dispensed at least 90 consecutive treatment days of antipsychotics during the measurement year.

Step 2: Determine the numerator. For each patient in the eligible population, by drug, identify all prescription events, start dates

and end dates. Identify the number of concurrent antipsychotic medication treatment events.
Step 2a: For each patient, identify the first day during the measurement year where the patient was being treated with two or more different antipsychotic medications; this is the concurrent antipsychotic medication treatment event start date.
Step 2b: Beginning with (and including) the start date, identify the number of consecutive days where the patient remains on two or more different antipsychotic medications. If the number of days =90 days, the patient is numerator compliant.
Step 2c: If the number of consecutive days on multiple antipsychotic medications is <90 days, identify the end date and identify the next day during the measurement year where the patient was being treated with two or more different antipsychotic medications. If the number of days between the end date and the next start date is =15 days, include the days in the concurrent antipsychotic medication treatment event (concurrent antipsychotic medication treatment events allow for a 15-day gap).
Step 2d: If the number of days between the end date and the next start date exceeds 15 days, end the event; using the new start date, continue to assess for concurrent antipsychotic medication treatment events.
Step 2e: Continue this process until the number of concurrent antipsychotic medication treatment days is =90 consecutive days (i.e., the patient is numerator compliant) or until the measurement year is exhausted (i.e., no concurrent antipsychotic medication treatment events were identified during the measurement year).

Step 3: Divide the numerator by the denominator to calculate the rate.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
This measure is part of the Healthcare Effectiveness Data and Information Set (HEDIS). As part of 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 Management Organizations and Preferred Provider Organizations via NCQA's online data submission system.

This measure has also been tested at the state level and could be reported by states if added to a relevant program.

S.25. **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.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Integrated Delivery System, Population : State

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Ambulatory Care : Clinician Office/Clinic, Behavioral Health/Psychiatric : Outpatient
If other:

S.28. **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

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
Multiple_Concurrent_Testing_10-12-15.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2799 (New Measure)
Measure Title: Use of Multiple Concurrent Antipsychotics in Children and Adolescents
Date of Submission: 10/9/2015
Type of Measure:

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

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

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

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

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion

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

2b4. For outcome measures and other measures when indicated (e.g., resource use):

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

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

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

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

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

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

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

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

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

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/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*).

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
  - Commercial health plans nationwide

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

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

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

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

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency

☒ health plan	☒ health plan
☒ other: State; Integrated Delivery System	☒ other: State; Integrated Delivery System

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*)
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 measure in Medicaid health plan data from one large mid-Atlantic state. In order to assess the measure's use for HEDIS, we conducted an additional analysis in commercial data from a large administrative database. Our samples were as follows.

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

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

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

Citations
Byrd VLH, Dodd AH. *Assessing the usability of encounter data for enrollees in comprehensive managed care across MAX 2007-2009*. December 2012 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*)
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 126,018 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 13,294, and across 49 commercial plans that had sufficient denominators (>30), the total was 11,895.

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

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

Mean	11,456
Minimum	1,545
25th	5,951
Median	10,393
75th	15,569
Maximum	24,161

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

Mean	783
Minimum	123
25th	319
Median	680
75th	976
Maximum	2,582

Denominator Size Distribution Across 49* Commerical Health Plans Nationwide (2012)

Mean	243
Minimum	31
25th	92
Median	168
75th	290
Maximum	1,566

* Of the 73 commercial plans included in the testing of this measure, 49 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.

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

Validity was demonstrated through a systematic assessment of face validity. Per NQF instructions we have described the composition of the technical expert panels which assessed face validity in the data sample questions above. In addition, validity was demonstrated through two types of analyses: correlations among measures using Spearman Correlation Coefficients (using commercial health plan data sample and Medicaid health plans 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 2b2.3.

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

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We assessed differences across multiple age strata (0-5, 6-11, 12-17, and total [0-17]), race/ethnicity (Hispanic; White, non-Hispanic; Black, non-Hispanic), and foster care status.

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*)

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

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

PLAN-LEVEL RELIABILITY

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

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

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

This measure achieved a reliability score above 0.7 for state-level reliability and about 0.7 for Medicaid health-plan level reliability. This measure achieved a higher level and narrower range of reliability in the state data compared to the health plan data.

	Average Reliability	Minimum Reliability

MAX States	.99	.96
Medicaid Health Plans	.64	.28
Commercial Health Plans	.42	.08

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?*)

As stated in 2a2.2, we estimated reliability with a beta-binomial model (Adams, 2009). A score of zero implies that all the variability in a measure is attributable to measurement error. A score of 1.0 implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one reporting entity from another. A score of 0.7 or higher indicates adequate reliability to distinguish performance between two entities and is considered acceptable. The testing results suggest that this measure has high reliability at the state level.

2b2. VALIDITY TESTING

2b2.1. 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*)

2b2.2. 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)*

Face Validity

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

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

Step 1: NCQA staff identifies areas of interest or gaps in care. Clinical expert panels (MAPs—whose members are authorities on clinical priorities for measurement) participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness and feasibility. This information is gathered into a work-up format. 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 NCQAs Board of Directors will be included in the next HEDIS year and reported as first-year measures.

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

Empirical Validity

As part of field testing, we assessed construct validity, which considers whether measures are capturing important aspects of a quality concept. We conducted two types of analyses: correlations among measures and rankings of health plans and states on measures.

We first tested for construct validity by exploring whether this measure was correlated with other related measures, including the *Follow-up Visit for Children and Adolescents on Antipsychotics* measure. The *Follow-up Visit* measure assesses the percentage of youth who have a follow-up visit with a prescriber within 30 days after the start of a new antipsychotic prescription. We hypothesized that organizations that perform well on one of these measures should perform well on the other measure. We calculated correlations using the Spearman correlation coefficients. 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 which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable.

We then explored whether entities that manage one aspect of antipsychotic prescribing for children and adolescents well, such as avoiding multiple concurrent antipsychotics, also manage other aspects of care well. This test shows if plans and states can be approximately ranked based on profiles of performance across multiple measures. Consistency of performance across measures suggests that the measures are assessing a dimension of quality. For state rankings, we compared the *Multiple Concurrent* measure with the *Use of Antipsychotics in Very Young Children* measure and the *Metabolic Monitoring for Children and Adolescents on Antipsychotics* measure. For the Medicaid health plan rankings we compared the Multiple Concurrent measure with the *Use of Higher than Recommended Doses of Antipsychotics in Children and Adolescents* measure and the *Use of First-Line Psychosocial Care for Children and Adolescents on Antipsychotics* measure. The *Antipsychotics in Very Young* measure assesses the percentage of youth under age six who were prescribed antipsychotics (a lower rate indicates better performance). The *Metabolic Monitoring* measure assesses the percentage of youth with ongoing antipsychotic use who had metabolic monitoring. The *High Dose* measure assesses the percentage of youth prescribed a higher-than-recommended dose of an antipsychotic. The

Psychosocial Care measure assesses the percentage of youth who receive first-line psychosocial care when newly prescribed an antipsychotic (among those youth that do not have a primary indication for an antipsychotic).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Face Validity Results

Step 1: This measure was developed to address inappropriate prescribing patterns for children and adolescents on antipsychotics. NCQA and five expert panels worked together in 2013 and 2014 to identify the most appropriate method for assessing the use of multiple concurrent antipsychotics among this patient population. Across the multiple expert panels that reviewed this measure, all panels concluded this measure was specified to assess multiple concurrent use of 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. This measure was rated a high priority by many commenters. Of 74 comments received, the majority (65 percent) supported it as-is or with suggested modifications. The CPM recommended moving this measure to first year data collection by a majority vote in May 2014.

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

Empirical Validity Results

Correlations

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

The results showed that among Medicaid health plans, there were no statistically significant correlations between the *Multiple Concurrent* measure and the other measures addressing antipsychotic use in children and adolescents. Among national commercial plans, there was moderate negative correlation between the *Follow-up Visit* and *Multiple Concurrent* measures (r=-.58, p=0.02). In addition to assessing correlations among the measures in this set, we examined correlations between performance on the measures and rates of hospitalization for mental health and substance use problems. However, we did not find consistent correlations.

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	Multiple Concurrent Antipsychotics[1]	Antipsychotics in Very Young Children[1]	Metabolic Monitoring[2]
1	5.7	**0.3**	14.2
2	6.6	**0.3**	19.4
3	**9.4**	**0.3**	20.6
4	7.7	0.2	**6.5**

5	3.3	**0.1**	**4.8**
6	**2.9**	**0.3**	18.7
7	**8.1**	0.2	20.0
8	7.1	**0.1**	14.8
9	7.7	**0.0**	**29.1**
10	4.1	**0.1**	19.6
11	**3.0**	**0.1**	**36.2**
Mean	**6.0**	**0.2**	**18.5**

[1]Lower rate indicates better performance
[2]Higher rate indicates better performance

Medicaid Health Plan Performance Rankings for One State

Plan	Multiple Concurrent Antipsychotics[1]	Higher than Recommended Doses[1]	First-Line Psychosocial Care[2]
3	3.8	**11.7**	41.7
9	**7.0**	8.3	48.6
6	**6.6**	4.9	30.1
17	3.3	**9.6**	**26.4**
2	5.1	4.4	**27.4**
8	4.6	5.4	43.5
4	3.3	5.8	46.9
5	3.9	4.9	42.4
1	5.6	5.6	51.6
11	5.1	5.7	43.8
16	3.3	4.0	56.6
15	6.3	5.7	28.0
12	4.3	4.7	43.3
13	4.5	**3.3**	30.7
7	**2.3**	4.6	**67.7**
14	4.6	5.4	64.3
10	**1.8**	**2.7**	**67.0**
Mean	**4.4**	**5.7**	**44.7**

[1]Lower rate indicates better performance
[2]Higher rate indicates better performance

2b2.4. 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?*)

Face Validity

The expert panels consulted showed good agreement that the measure as specified will accurately differentiate quality across states and health plans. Additionally this measure was rated as a high priority measure by the expert panels and by those who responded to the public comment. Our interpretation of these results is that this measure has sufficient face validity.

Empirical 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 confirmed our hypothesis that commercial plans that performed well on providing follow-up visits (higher rates indicate better performance) also performed well on avoiding multiple concurrent prescribing for those on antipsychotics (lower rates indicate better performance).

Ranking

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

2b3. EXCLUSIONS ANALYSIS

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

2b3.1. 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*)

2b3.2. 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*)

2b3.3. 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*)

2b4. 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 2b5.

2b4.1. 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

2b4.2. If an outcome or resource use 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.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

2b4.5. 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 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. 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*)

2b4.11. 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*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified

(*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR). 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.

2b5.2. 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*)

Variation in Performance Rates across MAX States (2008 data)

Mean Rate	10th	25th	50th	75th	90th	IQR
6.0	3.0	3.3	6.6	7.7	8.1	4.4

IQR: Interquartile range

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

Mean Rate	10th	25th	50th	75th	90th	IQR
4.4	2.9	3.3	4.5	5.4	6.4	2.1

IQR: Interquartile range

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

Mean Rate	10th	25th	50th	75th	90th	IQR
3.1	.7	1.4	3.1	4.6	5.1	3.2

IQR: Interquartile range

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The results indicate that there is 2.1% gap in performance between Medicaid plans at the 25th and 75th percentiles, a 3.2% gap in performance among commercial plans and a 4.4% gap in performance among states at the 25th and 75th percentiles. This means states at the 75th percentile have on average 504 more children and adolescents receiving multiple concurrent antipsychotics than states at the 25th percentile.

2b6. 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 SDS 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 SDS 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.

2b6.1. 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*)
N/A

2b6.2. 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*)
N/A

2b6.3. 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*)
N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)
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.

2b7.2. 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*)
N/A

2b7.3. 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*)
N/A

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)*

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.**

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.**

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. **Describe what you have learned/modified 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 a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Field testing results, more fully described in the Scientific Acceptability section, showed the measure is feasible to be collected by health plans and states using administrative claims data. Further, 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 managed care organization´s ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.

The HEDIS Compliance Audit addresses the following functions:
1) information practices and control procedures
2) sampling methods and procedures
3) data integrity

4) compliance with HEDIS specifications
5) analytic file production
6) reporting and documentation

In addition to the HEDIS Audit, NCQA provides a system to allow "real-time" feedback from measure users. Through our Policy Clarification Support System, NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures, including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation, information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the 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)*.
Broad public use and dissemination of these measures is encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Use by health care physicians in connection with their own practices is not commercial use. Commercial use of a measure requires the prior written consent of NCQA. As used herein, "commercial use" refers to any sale, license or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.

4. Usability and Use

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

4a. Accountability and Transparency

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

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

Planned	Current Use (for current use provide URL)
	Public Reporting Quality Compass http://www.ncqa.org/tabid/177/Default.aspx http://www.ncqa.org/tabid/836/Default.aspx The State of Health Care Quality Report Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Compass http://www.ncqa.org/tabid/177/Default.aspx The State of Health Care Quality Report http://www.ncqa.org/tabid/836/Default.aspx

4a.1. **For each CURRENT use, checked above, provide:**

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

QUALITY COMPASS: This measure has just been approved for use in Quality Compass for Medicaid, a tool that displays health plan-level performance rates for HEDIS measures. It is 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. The Quality Compass 2015 Medicaid tool includes data for 182 public reporting Medicaid health plan products, serving approximately 20 million covered lives. Benchmarks are calculated from a total pool of 244 public and non-public reporting health plan products, serving approximately 25 million covered lives.

THE STATE OF HEALTH CARE QUALITY REPORT: HEDIS measures are reported nationally and by geographic regions in the State of Health Care Quality Report, published by NCQA and summarizing findings on quality of care. In 2015 the report included measures on 15.4 million Medicare Advantage beneficiaries in 507 Medicare Advantage health plans, 103.9 million members in 413 commercial health plans, and 25.4 million Medicaid beneficiaries in 237 plans across 50 states.

4a.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

4a.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

4b. 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.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (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**

N/A

4b.2. **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.**
N/A

4c. 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).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No negative consequences have been reported since implementation.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are

compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
2337 : Antipsychotic Use in Children Under 5 Years Old

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

5a. Harmonization

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 completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
This new measure assesses multiple concurrent antipsychotic use children and adolescents who have continuous antipsychotic use. Measure 2337 is Antipsychotic Use in Children under 5 Years Old and assesses whether children under 5 are prescribed an antipsychotic at some point during the measurement year. Both measures are specified for the health plan level and use administrative claims as the data source. Both measures assess antipsychotic use; however, our measure has a broader age range (up to 18 years). In addition, the target population for this new measure is also focused only on those who have continuous use of antipsychotics for 90 days or more. In terms of measure focus, measure 2337 is focused on the utilization of antipsychotics among very young children for 30 days or more. The Use of Multiple Concurrent Antipsychotics in Children and Adolescents measure is focused on the receipt of multiple antipsychotics concurrently for at least 90 days during the measurement year. While both measures are assessing overuse/appropriateness of antipsychotics in children, what is being measured (or considered overuse) is different. While measure 2337 looks at any prescription for antipsychotics, our measure looks for a specific type of non-recommended practice (multiple concurrent use).

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 on 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 on 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
Bruce Bobbitt, PhD, LP, Optum
Peter Delany, PhD, LCSW-C, Substance Abuse and Mental Health Services Administration
Ben Druss, MD, MPH, Emory University
Frank A. Ghinassi, PhD, ABPP, Western Psychiatric Institute and University of Pittsburgh Medical Center
Rick Hermann, MD, Tufts Medical Center and UpToDate, Inc.
Connie Horgan, ScD, Brandeis University
Neil Korsen, MD, Maine Health
Charlotte Mullican, BSW, MPH, Agency for Healthcare Research and Quality
Harold Pincus, MD, Columbia University and RAND Corporation
Bruce L. Rollman, MD, MPH, University of Pittsburgh School of Medicine
Michael Schoenbaum, PhD, National Institute of Mental Health
John H. Straus, MD, Massachusetts Behavioral Health Partnership and Beacon Health Options

NCQA Committee on Performance Measurement (CPM)
Bruce Bagley, MD, American Academy of Family Physicians
Andrew Baskin, MD, Aetna
Patrick Conway, MD, MMSc, Center for Medicare & Medicaid Services
Jonathan D. Darer, MD, Geisinger Health System
Helen Darling, National Business Group on Health
Rebekah Gee, MD, MPH, FACOG, LSU School of Medicine and Public Health
Foster Gesten, MD, NYSDOH Office of Managed Care
David Grossman, MD, MPH, Group Health Physicians
Christine Hunter, MD (Co-Chair), US Office of Personnel Management
Jeffrey Kelman, MMSc, MD, Centers for Medicare & Medicaid Services
Bernadette Loftus, MD, The Permanente Medical Group
J. Brent Pawlecki, MD, MMM, The Goodyear Tire & Rubber Company
Susan Reinhard, RN, PhD, AARP
Eric C. Schneider, MD, MSc (Co-Chair), RAND Corporation
Marcus Thygeson, MD, MPH, Blue Shield of Califorina

NCQA Technical Measurement Advisory Panel
Andy Amster, MSPH, Kaiser Permanente
Kathryn Coltin, MPH, Independent Consultant
Lekisha Daniel-Robinson, Centers for Medicare and Medicaid Services
Marissa Finn, MBA, Cigna HealthCare
Scott Fox, MS, MEd, Independence Blue Cross
Carlos Hernandez, CenCalHealth
Kelly Isom, MA, RN, Aetna
Harmon Jordan, ScD, RTI International
Ernest Moy, MD, MPH, Agency for Healthcare Research and Quality
Patrick Roohan, New York State Department of Health
Lynne Rothney-Kozlak, MPH, Rothney-KozlakConsulting, LLC
Natan Szapiro, Independence Blue Cross

National Collaborative for Innovation in Quality Measurement (NCINQ) Measurement Advisory Panel
Mary Applegate, MD, Ohio Department of Job and Family Services
Katie Brookler, Colorado Department of Health Care Policy and Financing
Cathy Caldwell, MPH, Alabama Department of Public Health
Jennifer Havens, MD, NYU School of Medicine
Ted Ganiats, MD, University of California, San Diego
Darcy Gruttadaro, JD, National Allegiance on Mental Illness
Virginia Moyer, MD, MPH, FAAP, Baylor College of Medicine, USPSTF
Edward Schor, MD, Lucile Packard Foundation for Children's Health
Xavier Sevilla, MD, FAAP, Whole Child Pediatrics
Gwen Smith, Illinois Department of Healthcare and Family Services/Health Management Associates
Janet (Jessie) Sullivan, MD, Hudson Health Plan
Kalahn Taylor-Clark, PhD, MPH, George Mason University
Craig Thiele, MD, CareSource
Charles Wibbelsman, MD, Kaiser Permanente Medical Group, Inc.
Jeb Weisman, PhD, Children's Health Fund

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:**

Ad.4 **What is your frequency for review/update of this measure?** Every 3 years

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

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