

## 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 #:** 3313

**Corresponding Measures:**

**Measure Title:** Follow-Up Care for Adult Medicaid Beneficiaries Who are Newly Prescribed an Antipsychotic Medication

**Measure Steward:** Centers for Medicare & Medicaid Services

**Brief Description of Measure:** Percentage of new antipsychotic prescriptions for Medicaid beneficiaries aged 18 years and older who have completed a follow-up visit with a provider with prescribing authority within four weeks (28 days) of prescription of an antipsychotic medication.

**Developer Rationale:** Among individuals with serious mental illness, physical health problems such as cardiovascular disease, metabolic disorders, and infectious disease are more prevalent compared to the general population. Antipsychotic medications can exacerbate existing physical problems as well as increase a patient's risk for developing new health concerns such as metabolic complications. Timely follow-up with a provider following the prescription of antipsychotic medications is an essential first step to ensure that physical impacts of antipsychotic medications are identified and addressed early. Early follow up is also critical to monitor for treatment effectiveness and modify dosage as necessary, as well as to identify and address any barriers to treatment adherence. By proactively following up with patients who are prescribed antipsychotic medications, providers can identify problems early in the course of treatment and minimize potential harms associated with use of those medications. Regardless of the care setting in which a patient is being treated, comprehensive assessment of both physical and mental health factors is an essential aspect of treatment with antipsychotic medications.

**Numerator Statement:** The percentage of Medicaid beneficiaries aged 18 years and older with new antipsychotic prescriptions who completed an outpatient follow-up visit with a provider with prescribing authority within 28 days of the new antipsychotic prescription fill.

**Denominator Statement:** New antipsychotic prescriptions for Medicaid beneficiaries aged 18 years and older.

**Denominator Exclusions:** Medicaid beneficiaries with an acute inpatient admission during the 28-day follow-up period after prescription of an antipsychotic medication

- Patients who expired within 28 days of new antipsychotic prescription fill date.

**Measure Type:** Process

**Data Source:** Claims

**Level of Analysis:** Population: Regional and State

**IF Endorsement Maintenance – Original Endorsement Date:** 05/16/2018

**Most Recent Endorsement Date:** 05/16/2018

## Preliminary Analysis: Maintenance of Endorsement

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

### Criteria 1: Importance to Measure and Report

#### 1a. [Evidence](#)

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

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

**The developer provides the following description for this measure:**

- This is a maintenance process measure at the population level that assesses follow-Up Care for Adult Medicaid Beneficiaries Who are Newly Prescribed an Antipsychotic Medication.
- The developer provides a [logic model](#) that depicts the prescription of a new antipsychotic medication, leading to the completion of a follow-up visit to monitor treatment effectiveness, leading to the evaluation of any side effects/treatment effectiveness/dosage modification/barriers to medication adherence, leading to reduced medication side effects/improved effectiveness of treatment/improved medication adherence. By proactively following up with patients who are prescribed antipsychotic medications, providers can identify problems early in the course of treatment and minimize potential harms associated with those medications.

**The developer provides the following evidence for this measure:**

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

### Summary of prior review in 2016

- The developer presented four clinical guidelines in support of follow-up for individuals with new antipsychotic prescriptions. The Standing Committee agreed that any type of health monitoring and follow-up is important for this target population
- The Standing Committee discussed the importance of including telemedicine as a follow-up method for the measure to improve access. The current specifications include telephone follow-up, and the developer stated they intend to include telemedicine codes in future specifications.
- The Standing Committee expressed concerns that Medicaid claims do not identify the content of the follow-up visit and that a follow-up encounter may not be specific to antipsychotic use.

### Changes to evidence from last review

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

The developer provided updated evidence for this measure:

- The developer conducted an environmental scan to identify any new evidence pertaining to the measure, as well as interviews with key stakeholders and subject matter experts.
- The developer states that Guideline two, from 2004, was replaced with an update to the recommendations, from 2020. The recommendation comes from a systematic review that includes 167 double blind randomized trials, 465 randomized clinical trials and 402 randomized placebo-controlled trials ranging from 1960 to 2018.
  - “Patients may take between 2 and 4 weeks to show an initial response and longer periods of time to show full or optimal response. Monitoring of the patient’s clinical status for 2-4 weeks is warranted on a therapeutic dose unless the patient is having uncomfortable side effects.” (Grade A1: High confidence that the evidence reflects the true effect and that the benefit outweighs the harm.) The recommendation comes from a systematic review that includes 167 double blind randomized trials, 465 randomized clinical trials and 402 randomized placebo-controlled trials ranging from 1960 to 2018).

### Questions for the Committee:

- *The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?*
- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

### Guidance from the Evidence Algorithm

Process measure (Box 1) -> Evidence is based on a guideline and systematic review of a body of evidence (Box 3) -> Quantity provided but not quality or consistency (Box 4) -> Grade of recommendation indicates high quality evidence (Box 6) -> Rate as MODERATE

Preliminary rating for evidence:  High  Moderate  Low  Insufficient

---

1b. [Gap in Care/Opportunity for Improvement](#) and [Disparities](#)

---

**Maintenance measures – increased emphasis on gap and variation**

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

- The unit of analysis is at the state level. Data were provided for 13 states between January 1, 2018 and December 31, 2018.
  - Mean performance rate was 46.72 percent with a standard deviation of 6.04 percent.
  - The lowest state had follow up visits with 35.86 percent of patients newly prescribed an antipsychotic and the highest rate was 58.72 percent.
  - The 25th percentile was 43.62 percent, median was 45.17 percent and the 75th percentile was 50.68 percent.

**Disparities**

- Disparities data were also provided for 13 states between January 1, 2018 and December 31, 2018. Results were stratified by age, race, ethnicity, dual eligibility status, gender and Medicaid eligibility category.
  - Younger patients and those age 75+ were less likely to have a follow up visit than those aged 35-74 years old.
  - White non-Hispanic (47.9%) and Hispanic (47.3%) patients were more likely to have a follow up visit than black non-Hispanic (40.4%) patients.
  - Dual eligible (48.5%) patients were more likely to have a follow up than Medicaid only (44.2%) patients.
  - Females (48.9%) were more likely to have a follow up visit than male (41.2%) patients.
  - CHIP patients (38.5%) were less likely to have a follow up visit than aged (46.4%), disabled (45%) or adult (47.5%) patients. However, CHIP patients are a very small part of the overall sample, likely reflecting pregnant women and 18 year olds.

**Questions for the Committee:**

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

Preliminary rating for opportunity for improvement:  High  Moderate  Low   
Insufficient

**Committee Pre-evaluation Comments:**

**1a. Evidence**

- process measure, evidence based on guidelines. No data presented that adherence to this measure is improving over time.
- There is evidence to support the measure.
- The developer completed a systematic review, demonstrating a strong body of evidence. The evidence demonstrates the need for follow up for patients who start antipsychotic medications. Follow up can

support the prevention and management of other chronic physical conditions. Not aware of any new studies, this review of the evidence is comprehensive.

- Given the strength of the evidence, including updated evidence, further discussion and vote on evidence does not seem necessary.
- Follow-up appointments are important to assess whether an antipsychotic medication is working or if changes to the treatment regimen should be made. A 28 day timeframe is appropriate. However, there is a great deal of nonadherence to psychiatric medication which influences outcomes and a process measure will not adequately capture this.
- Evidence is reasonable
- This is a maintenance process measure-new evidence was incorporated.- no change
- Evidence remains reasonable

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

- gap in care is based solely on state agency reported from 13 states (not identified) during calendar year of 2018. average rate reported, rates within percentiles reported: 25th%tile: 43.6% and 75th %tile is about 51%
- Several notable gaps that warrant a national performance measure.
- At the state level, there is a range in performance with regard to follow ups after newly prescribed antipsychotic medications. There are also disparities with regard to age, race, insurance, gender and Medicaid eligibility.
- Median performance rate is below 50% suggesting ample room for improvement; some disparities noted by age, race, ethnicity, dual eligibility status, gender and Medicaid eligibility category.
- There is a gap in care for follow-up care. However, this may be due to a lack of psychiatric providers, and therefore the measure may not improve much.
- The gap is real, meaningful and deserves work to improve it.
- The performance gap is at the individual state level. The evidence does identify the disparities through population subgroups. Younger patients and 75 and up patients were less likely to have a follow up appointment. Outreach efforts should increase with those populations.
- Still an important gap, not topped out

## Criteria 2: Scientific Acceptability of Measure Properties

Complex measure evaluated by Scientific Methods Panel?  Yes  No

Evaluators: [Staff](#)

---

### 2a. Reliability: [Specifications](#) and [Testing](#)

---

**For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.**

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

**For maintenance measures – less emphasis if no new testing data provided.**

**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.

### Specifications:

- Measure specifications are clear and precise.

### Reliability Testing:

- Reliability testing conducted at the Accountable Entity Level:
  - The developer examined signal to noise ratio using the beta-binomial model (Adams, 2009). Thirteen states, representing a cross-section of U.S. regions, were used for measure testing using Medicaid and Medicare data from 9/1/2017-12/31/2017.
    - The average signal to noise ratio was .991 and ranged from .977 to .999 across the states.
    - These results are similar to the findings from the prior 2016 submission, but reflect updated 2017-2018 claims data.
    - The developer states that this high level of reliability means the measure can be used to distinguish between states with respect to healthcare quality. The high level of reliability is likely supported by the large sample size at the state level.

### Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are measure specifications adequate)?*

Preliminary rating for reliability:  High  Moderate  Low  Insufficient

---

2b. Validity: [Validity testing](#); [Exclusions](#); [Risk-Adjustment](#); [Meaningful Differences](#); [Comparability](#); [Missing Data](#)

---

**For maintenance measures – less emphasis if no new testing data provided.**

**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.

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

### Validity Testing

- Validity testing conducted at the Accountable Entity Level:
  - The developer examined validity through correlations between NQF 3313 and two HEDIS measures: 1) Follow-Up after Hospitalization (seven and 30 days) for Mental Illness and 2) Follow-Up Care for Children Prescribed ADHD Medication upon initiation and continuation and maintenance phases.
  - They hypothesized that there will be a positive correlation with both measures because the processes assessed in these measures are related or similar to the antipsychotic follow up measure. Of the 13 states, only nine had the HEDIS measures available. HEDIS data came from measurement year 2018. Spearman correlations were calculated for these nine states.
    - NQF 3313 had a moderately positive correlation with seven and 30 day follow up after hospitalization for mental illness ( $r=0.42$  and  $r=0.38$  respectively). Results were not statistically significant, likely due to the small sample size ( $n=9$ )

- NQF 3313 had a moderately positive correlation with the HEDIS measure for ADHD medication initiation for the initiation and continuation phase ( $r=0.58$  and  $=-0.53$ , respectively). Results were not statistically significant, likely due to the small sample size ( $n=9$ )
- In 2016, the developer conducted face validity with a technical expert panel (TEP). The TEP reviewed the measure and nearly 75 percent of respondents agreed that the measure's performance scores can distinguish between good quality and poor quality of care for Medicaid beneficiaries.

### **Exclusions**

- The measure has the following exclusions:
  - Patients who were hospitalized during the four-week follow-up period after initial prescription of antipsychotic medication
  - Patients who died during the four-week follow-up period after initial prescription of antipsychotic medication.
- The developer tested the exclusion for each measure separately and together. Together, these exclusions totaled 5.3 percent of the eligible population. The hospitalization exclusion accounted for 4.8 percentage points and the death exclusion accounted for the remaining 0.5 percentage points. The overall follow up rate with no exclusions is 49.0 percent and with exclusions it falls to 48.8 percent.

### **Risk-Adjustment**

- The measure is not risk adjusted or stratified.

### **Meaningful Differences**

- Seven of the 13 states had a performance rate and 95 percent confidence interval below the overall average for the 13 states. Nine of the 13 states had a performance rate below 50 percent, suggesting no problem with the measure topping out.
- A t-test comparing a randomly selected state above the 75th percentile performance and a randomly selected state below the 25th percentile performance was statistically significant ( $p<.001$ ). The Interquartile range was 7.1 percentage points.

### **Missing Data**

- States were selected based on having sufficient data for inclusion in the analyses. They also examined the Transformed Medicaid Statistical Information System (T-MSIS) Research Identifiable Files (RIF) 2017 data and CY18 to identify states with minimal missing data for the key variables: 1) National Drug Codes (NDCs) and fill dates; and 2) procedure and diagnosis codes in the outpatient and inpatient claims files.
- From 2016 submission: The only data element with more than 5 percent missing data is the Current Procedural Terminology (CPT)/Healthcare Common Procedure Coding System (HCPCS) code. Medicaid eligibility was missing for up to 4 percent of beneficiaries in one state. Race was missing for a large proportion of enrollees (range 0-43.8%).

### **Comparability**

- The measure only uses one set of specifications for this measure.

**Questions for the Committee regarding validity:**

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- Does the lack of statistically significant correlations, likely due to small sample size, between the measure and outcome measures cause any concern?

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

**Committee Pre-evaluation Comments:**

**2a. Reliability**

- 2a1. Reliability-Specifications
  - specifications are clear for use with Medicaid claims data
  - No concerns
  - Data elements and specifications are clearly defined. Signal to noise ration high. High reliability.
  - No concerns.
  - Are telehealth appointments considered completed appointments?
  - good
  - No concerns, clearly defined.
  - has been reliable
- 2a2. Reliability – Testing
  - Given aggregated data at state level for only 13 states, the team is left with beta binomial model to statistically explore capacity to distinguish higher vs. lower performing states.
  - No concerns
  - No.
  - No concerns.
  - No. This is a claims based measure and will be more reliable than not.
  - no
  - No concerns.
  - reliability testing is adequate

**2b. Validity**

- validity explored by examining correlation with adherence rates for two HEDIS measures and found only moderate correlations. The face validity was also explored using a survey item from a "technical expert panel" in 2016. no data presented that adherence is associated with improved outcomes.
- No concerns
- No, several validity tests conducted.
- No. While lack of statistically correlation is notable, it is likely due to small sample size. Reassured by qualitative input from experts suggesting adequate validity.
- No
- no concerns
- No concerns.



- Validity testing appropriate and sound

## **2b2-2b6. Potential threats to validity**

- 2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)
  - risk adjustment may not be feasible with this data source
  - No concerns.
  - Exclusions are appropriate. Could discuss with the committee if excluding hospitalized patients is necessary. Could be that they are lost to f/u as a result of hospitalization. Measure not risk adjusted or stratified.
  - Exclusions are clinically appropriate; no risk adjustment.
  - Variability in access to care and quality of care exists, so measure should be risk-adjusted.
  - The logic model highlights how folks with serious mental illnesses and who need antipsychotics are at high risk for medical illnesses and early death and hence deserve follow-up to check on these issues as well as how they are doing on a specific dose of antipsychotic medication. Given this it makes NO sense to exclude individuals who are admitted to an inpatient unit or die during the measurement period. It is far better to measure and report such instances (categorized by inpatient admissions and also by death) so we can learn how to better predict and avoid these sequelae.
  - The measure does use exclusions. The measure is not risk adjusted. Performance measures were high.
  - OK
- 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)
  - unable to assess
  - None.
  - Yes, there are meaningful differences between states. Only one set of specifications. Cannot tell the impact of missing data.
  - Comparison between states suggests meaningful differences about quality. No comparability issues. Missing data hard to assess as states were selected based on having sufficient data for inclusion in the analyses.
  - No
  - No significant issues
  - Missing data was accounted for and states were chosen with sufficient data.
  - No

## **Criterion 3. [Feasibility](#)**

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

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

- All data elements are available in Medicaid claims (TMSIS) and Medicare Part A, B, C and D administrative claims.
- The developer used the Data Quality Atlas to identify states with sufficient data quality.
- No applicable fees or licensing required.

### ***Questions for the Committee:***

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?

Preliminary rating for feasibility:  High  Moderate  Low  Insufficient

## Committee Pre-evaluation Comments:

### 3. Feasibility

- the main limitation is use of Medicaid claims data--this measure uses adjudicated pharmacy claims data and encounter data..
- No issues with feasibility.
- Data come from electronic claims and coded by someone other than person collecting information. No issues with the data collection strategy.
- No concerns. Data are routinely generated and used and available.
- This is a claims-based measure and data will be available to measure.
- feasibility is good
- Implementation may be an issue.
- Feasible

## Criterion 4: Use and Usability

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

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

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

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

#### Current uses of the measure

Publicly reported?  Yes  No

Current use in an accountability program?  Yes  No  UNCLEAR

Planned use in an accountability program?  Yes  No  NA

#### Accountability program details

- The Centers for Medicare & Medicaid Services (CMS) have approved NQF #3313 for states to use to assess the continuum of mental health care for antipsychotic use as part of the Section 1115 Demonstration for Serious Mental Illness (SMI) and Serious Emotional Disturbance (SED).
  - The developer reports that seven states will likely report NQF #3313 in the near future: District of Columbia, Idaho, Indiana, Oklahoma, Utah, Vermont and Washington.

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

**Feedback on the measure by those being measured or others**

- The developer provided assistance with measure implementation, including steps to calculate measure performance, during a CMS-sponsored webinar on June, 20 2019.
- No feedback has been received on measure performance or implementation from measured entities.

**Questions for the Committee:**

- *Are there plans for this measure to be publicly reported or used in an accountability program?*
- *How has the measure been vetted in real-world settings by those being measured or others?*

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

---

**4b. Usability (4a1. [Improvement](#); 4a2. [Benefits of measure](#))**

---

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

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

**Improvement results**

- State-level performance rates ranged from 35.9 percent to 58.7 percent.
- For nine of the 13 states included in this analysis, less than half of Medicaid beneficiaries who received a new antipsychotic medication had an outpatient follow-up visit within 28 days.
- Additionally, state-level performance in 2014 and 2018 was similar for the two states used in both measure testing submissions (3.2 percentage point improvement for one state and -0.5 percentage point decline for the other state).

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

**Unexpected findings (positive or negative) during implementation**

- There have been no unexpected findings from implementation of this measure.

**Potential harms**

- No harms have been identified.

**Questions for the Committee:**

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

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

Preliminary rating for Usability:  High  Moderate  Low  Insufficient

## Committee Pre-evaluation Comments:

### 4a. Use

- not publicly reported, no feedback
- No concerns.
- Not being publicly reported. Results are not available outside of the groups being measured. No accountability applications at the moment. Planned use for accountability. Measure users are provided with instructions and assistance, but no feedback has been collected.
- While data are not currently used in accountability programs, CMS has approved its usage for 1115 waivers for numerous states. Would like to see more opportunities for bi-directional feedback.
- Yes
- It's NOT being used at this time and the developer states that 7 states will likely be using it in the near future. I think we might consider trusting but verifying. I mean either provisionally accepting it but asking the developer to come back in 1-2 years with use on data or alternately ask them to come back in 1 year and approve it then if it's being used.
- No feedback has been received on the measure. States to use to assess the continuum of mental health care for antipsychotic use as part of the Section 1115 Demonstration for Serious Mental Illness (SMI) and Serious Emotional Disturbance (SED).
- Not in broad use yet.

### 4b. Usability

- why only 13 states for the data source? Can this be used by all states? This is important if BH measures in Core Set are to be reported by 2024ar
- No unexpected harms identified.
- There is a wide range of follow up rates by state. If states participate and get more data/feedback, follow up rates could improve and thus improve patient outcomes. There are no unexpected findings and no harms.
- No harms/unintended consequences, whereas many benefits toward improving quality of care.
- Benefits outweigh the risks.
- No evidence about this at this time
- No harms; Benefit-By proactively following up with patients who are prescribed antipsychotic medications, providers can identify problems early in the course of treatment and minimize potential harms associated with use of those medications.
- Could impact patient care positively.

## Criterion 5: [Related and Competing Measures](#)

### Related measures

- NQF #0108 *Follow-Up Care for Children Prescribed ADHD Medication (ADD)*
- NQF #3539e *Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting*

### Harmonization

- The developer has harmonized the measures to the extent possible.
- Measured population: NQF 0108 measures children on ADHD medications. NQF 3539e measures individuals over 65.
- Follow-up time period: Similar to NQF #3313, NQF 0108 uses the same follow-up period of “within four weeks (28 days) of prescription.”
- Setting: NQF 3539e measures individuals in an inpatient setting.

## Committee Pre-evaluation Comments:

### 5: Related and Competing Measures

- All related measures have been harmonized.
- Two related measures with appropriate harmonization.
- Should make sure the denominators for this measure and the antipsychotic adherence measure are harmonized. Would also be interesting to compare results to see if higher follow-up rates increased adherence.
- The ADHD measure is not a true competitor and the majority of factors one works on on an inpatient unit to successfully encourage continuous treatment are different than in the outpatient sphere.
- No competing measures; related-NQF #0108 Follow-Up Care for Children Prescribed ADHD Medication (ADD) • NQF #3539e Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting
- Not significant

## Public and NQF Member Comments (Submitted as of Month Day, Year)

### Member Expression of Support

- No public comments received.

### Scientific Acceptability Evaluation

#### RELIABILITY: SPECIFICATIONS

1. Have measure specifications changed since the last review?  Yes  No
2. Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?  Yes  No

**Submission document:** Items sp.01-sp.30

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

3. Briefly summarize any changes to the measure specifications and/or concerns about the measure specifications.
  - The developer attests that there have been no changes to the measure specifications since last review.

#### RELIABILITY: TESTING

4. Did the developer conduct new reliability testing?  Yes  No
  - 4a. If no, summarize the Standing Committee’s previous feedback:

**4b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee’s feedback from the previous review:**

- In 2016, the developer used Medicare enrollment data from the Medicare Beneficiary Summary File and from several sources of claims data in CMS’s Data Extract System. In 2021, the developer used data from the Transformed Medicaid Statistical Information System (T-MSIS) which contains beneficiary, service utilization, administrative claims, and expenditure data from the Medicaid population, including both fee-for-service and managed care payers.
- In both 2021 and 2016, the developer used a beta-binomial model to indicate signal-to-noise.
- 2016 testing found an average signal-to-noise value of 0.98, ranging from 0.88 to 0.99 across states (n=15 states).

**Reliability testing level:**  **Accountable-Entity Level**  **Patient/Encounter Level**  **Neither**

**5. Reliability testing was conducted with the data source and level of analysis indicated for this measure:**

**Yes**  **No**

**6. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was empirical VALIDITY testing of patient-level data conducted?**

**Yes**  **No**

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

- Signal to noise reliability for this measure was tested using a beta-binomial model to assess the level of confidence as it relates to the performance of one reported entity from another.
- The developer also calculated signal-to-noise reliability for each state individually.

**Submission document:** Question 2a.10

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

- The average signal-to-noise reliability was 0.991 with a range of 0.977 to 0.999 across states.
- The developer states that this suggests the measure was highly reliable in distinguishing performance across states.
- Among the state-level reliability estimates, no individual state’s reliability was lower than 0.979.
  - The developer states that their high reliability is likely supported by large sample sizes at the state level. The average number of newly filled antipsychotic prescriptions per state was 21,928 (ranging from 3,176 to 104,074), and the average number of prescriptions with continuity treatment per state was 10,047 (ranging from 1,647 to 46,260).

**9. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.**

**Submission document:** Question 2a.10-12

**Yes**

**No**

**Not applicable**

**10. Was the method described and appropriate for assessing the reliability of ALL critical data elements?**

**Submission document:** Question 2a.10-12

**Yes**

No

Not applicable (patient/encounter level testing was not performed)

11. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

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

**Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity level testing has not been conducted)

**Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

**Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

12. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

- Specifications are precise (Box 1) -> Empirical testing using the measure as specified (Box 2) -> Testing conducted at the accountable entity level (Box 4) -> Testing method is appropriate (signal-to-noise) (Box 5) -> Based on the reliability statistic and scope of testing, there is high certainty the measure is reliable (Box 6a) -> Rate as HIGH

#### VALIDITY: TESTING

13. **Did the developer conduct new validity testing?**  Yes  No

**14a. If no, summarize the Standing Committee's previous feedback:**

- [Summary]

**14b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:**

- In 2016, the developer conducted face validity. Out of an 11-member TEP, eight agreed that the measure distinguishes good quality from poor quality of care. Two were neutral and one disagreed with the statement.
- In 2021, the developer performed convergent validity testing to determine if the measure was positive correlated with two HEDIS measures (described below).

14. **Validity testing level (check all that apply):**

**Accountable-Entity Level**  **Patient or Encounter-Level**  **Both**

**NOTE:** Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

15. **If patient/encounter level validity testing was provided, was the method described and appropriate for assessing the accuracy of ALL critical data elements?** **NOTE:** Data element validation from the literature is acceptable.

Yes

No

**Not applicable** (patient/encounter level testing was not performed)

16. **Method of establishing validity at the accountable-entity level:**

**NOTE:** Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

**Submission document:** Questions 2b.01-02

**Face validity**

**Empirical validity testing at the accountable-entity level**

N/A (accountable-entity level testing not conducted)

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

**Submission document:** Question 2b.02

Yes

No

**Not applicable** (accountable-entity level testing was not performed)

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

**Submission document:** Question 2b.02

- The developer assessed whether NQF #3313 was correlated with two other similarly-calculated measures, HEDIS *Follow-Up After Hospitalization for Mental Illness (FUH)* and HEDIS *Follow-Up Care for Children Prescribed ADHD Medication (ADD)*.
- The developer calculated the Spearman correlation coefficient using state-level data derived from the Medicaid T-MSIS TAF RIFs as well as HEDIS data. HEDIS data was only available for 9 of the 13 states in the testing cohort. Due to the small sample size of 9 states, Spearman was calculated rather than Pearson correlations.

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

**Submission document:** Questions 2b.03-04

- NQF #3313 showed a moderately positive correlation with the HEDIS FUH measure for the 7-Day and 30-day indicators for follow-up after hospitalization for mental illness ( $r=0.42$  and  $r=0.38$ , respectively).
- NQF #3313 also showed a moderately positive correlation with the HEDIS ADD measure for the “initiation” and “continuation and maintenance” phase indicators ( $r=0.58$  and  $r=0.53$ , respectively).
  - The developer states that due to the small sample size ( $n=9$  states), they were not surprised that convergent validity results were not statistically significant at the level of  $p<0.05$ .
  - The developer also notes that differences in measure intent, as well as target population, may help explain the moderate strength in correlation. The FUH measure focuses on ensuring care coordination for the mental health population from the inpatient setting to the outpatient setting, rather than outpatient only, and the ADD measure is focused on medication management for a child-focused, rather than adult-focused, population.

**VALIDITY: ASSESSMENT OF THREATS TO VALIDITY**

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

**Submission document:** Questions 2b.15-18.

- Exclusion testing was performed as part of the 2016 submission. The developer notes that exclusions have limited impact on denominator size and measure performance, but that they were maintained the exclusions in the measure specifications, given that (1) the clinical advisory workgroup contributed to the development of these exclusions and the TEP supported the



exclusions as appropriate during their review of testing results, and (2) the exclusions are relatively straightforward to calculate and are not expected to impose significant additional burden in measure implementation

**21. Risk Adjustment**

**Submission Document:** Questions 2b.19-32

**22a. Risk-adjustment method**

- None       Statistical model       Stratification  
 Other method assessing risk factors (please specify)

**22b. If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

- Yes       No       Not applicable

**22c. Social risk adjustment:**

22c.1 Are social risk factors included in risk model?       Yes       No       Not applicable

22c.2 Conceptual rationale for social risk factors included?       Yes       No

22c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus?       Yes       No

**22d. Risk adjustment summary:**

22d.1 All of the risk-adjustment variables present at the start of care?       Yes       No

22d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion?  
 Yes       No

22d.3 Is the risk adjustment approach appropriately developed and assessed?       Yes       No

22d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration)  
 Yes       No

22d.5. Appropriate risk-adjustment strategy included in the measure?       Yes       No

**22e. Assess the risk-adjustment approach**

- [Summary]

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

**Submission document:** Questions 2b.05-07

- State-level performance rates ranged from 35.9 percent to 58.7 percent, with an average of 45.8 percent and a standard deviation of 5.0 percent. Seven of the thirteen states had performance rates and 95 percent confidence intervals below the overall average.
- The developer performed t-tests to compare a randomly selected state above the 75<sup>th</sup> percentile and a randomly selected state below the 25<sup>th</sup> percentile. The difference was statistically significant at  $p < 0.001$ , indicating this is likely a true difference in performance.

*For cost/resource use measures, does this measure identify meaningful differences about cost and resource use between the measured entities?*

**23. Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

**Submission document:** Questions 2b.11-14.

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

- States were selected based on having sufficient data for inclusion in the analyses. They also examined the Transformed Medicaid Statistical Information System (T-MSIS) Research Identifiable Files (RIF) 2017 data and CY18 to identify states with minimal missing data for the key variables: 1)

National Drug Codes (NDCs) and fill dates; and 2) procedure and diagnosis codes in the outpatient and inpatient claims files.

- From 2016 submission: The only data element with more than 5 percent missing data is the Current Procedural Terminology (CPT)/Healthcare Common Procedure Coding System (HCPCS) code. Medicaid eligibility was missing for up to 4 percent of beneficiaries in one state. Race was missing for a large proportion of enrollees (range 0-43.8%).

#### For cost/resource use measures ONLY:

*If not cost/resource use measure, please skip to question 25.*

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

Yes    Somewhat    No (If “Somewhat” or “No”, please explain)

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

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

- High** (NOTE: Can be HIGH only if accountable-entity level testing has been conducted)
- Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity level testing has NOT been conducted)
- Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)
- Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the accountable-entity level and the patient/encounter level is required; if not conducted, should rate as INSUFFICIENT.)

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

Threats to validity were addressed (Box 1) -> Empirical testing was of the measure as specified (Box 2) -> Empirical testing conducted at the accountable entity level (Box 5) -> Method was appropriate for assessing hypothesized relationships (Box 6) -> Based on testing results there is moderate certainty that the measures is valid -> Rate as MODERATE

#### FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

29. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

- High
- Moderate
- Low
- Insufficient

30. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

- [Summary]

#### ADDITIONAL RECOMMENDATIONS

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

- [Summary]

## Developer Submission

---

### Criteria 1: Importance to Measure and Report

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

---

**1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.**

**[Response Begins]**

Yes

**[Yes Please Explain]**

Guideline two, from 2004, was replaced with an update to the recommendations, from 2020.

**[Response Ends]**

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

**2021 Submission:**

Updated evidence information here.

**2018 Submission:**

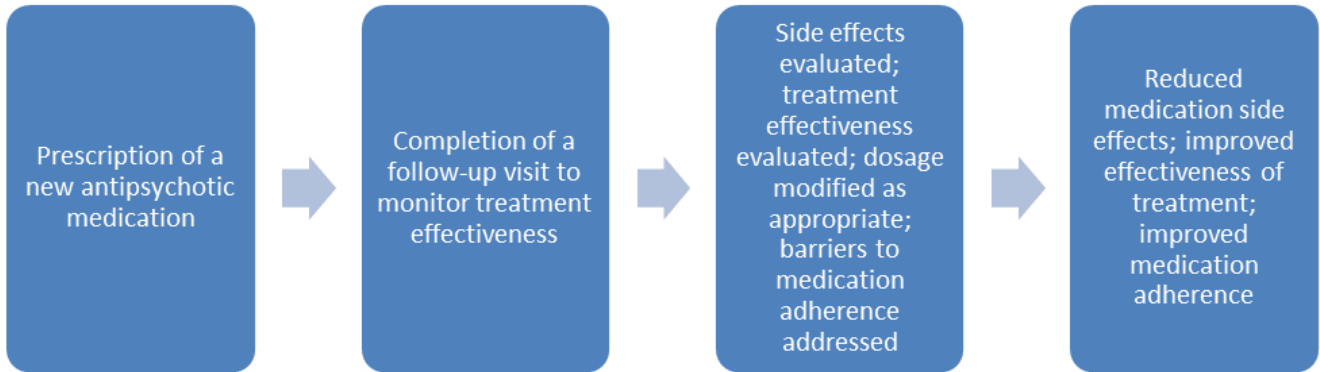
Evidence from the previous submission here.

**1a.01. Provide a logic model.**

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

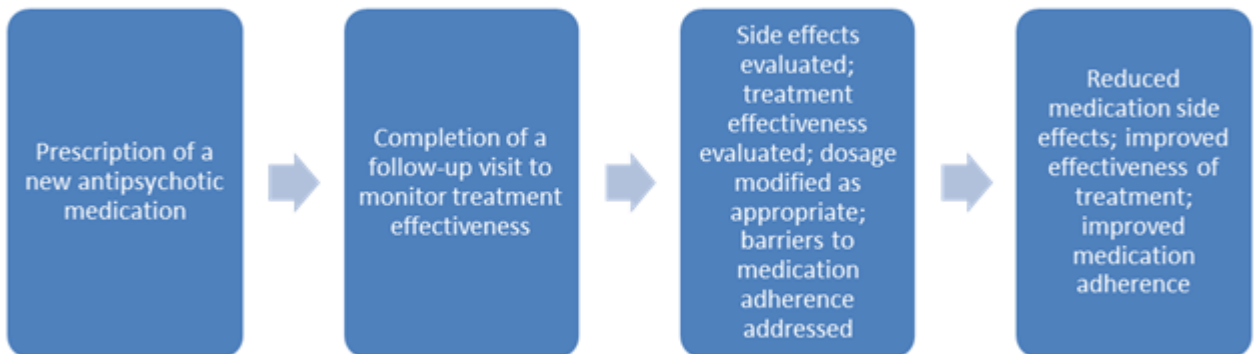
**[Response Begins]**

**2022 Submission**



**2016 Submission**

Among individuals with serious mental illness, physical health problems such as cardiovascular disease, metabolic disorders, and infectious disease are more prevalent compared to the general population. Antipsychotic medications can exacerbate existing physical problems as well as increase a patient’s risk for developing new health concerns such as metabolic complications. Timely follow-up with a provider following the prescription of antipsychotic medications is an essential first step to ensure that physical impacts of antipsychotic medications are identified and addressed early. Early follow up is also critical to monitor for treatment effectiveness and modify dosage as necessary, as well as to identify and address any barriers to treatment adherence. By proactively following up with patients who are prescribed antipsychotic medications, providers can identify problems early in the course of treatment and minimize potential harms associated with use of those medications. Regardless of the care setting in which a patient is being treated, comprehensive assessment of both physical and mental health factors is an essential aspect of treatment with antipsychotic medications.



[Response Ends]

**1a.02. Select the type of source for the systematic review of the body of evidence that supports the performance measure.**

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

**[Response Begins]**

Clinical Practice Guideline recommendation (with evidence review)

**[Response Ends]**

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, add additional tables by clicking “Add” after the final question in the group.

**Evidence - Systematic Reviews Table (Repeatable)**

Group 1 - Evidence - Systematic Reviews Table

**1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.**

**[Response Begins]**

**2022 and 2016 Submissions**

American Diabetes Association (ADA), American Psychiatric Association (APA), American Association of Clinical Endocrinologists (AAACE), North American Association for the Study of Obesity (NAASO). “Consensus development conference on antipsychotic drugs and obesity and diabetes.” *Diabetes Care*. 2004;27(2): 596-601. Available at <http://care.diabetesjournals.org/content/27/2/596>

**[Response Ends]**

**1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.**

**[Response Begins]**

**2022 and 2016 Submissions**

*Follow-up monitoring:*

“The patient’s weight should be reassessed at 4, 8, and 12 weeks after initiating or changing second generation antipsychotic (SGA) therapy and quarterly thereafter at the time of routine visits.

“Fasting plasma glucose, lipid levels, and blood pressure should also be assessed 3 months after initiation of antipsychotic medications. Thereafter, blood pressure and plasma glucose values should be obtained annually or more frequently in those who have a higher baseline risk for the development of diabetes or hypertension. In those with a normal lipid profile, repeat testing should be performed at 5-year intervals or more frequently if clinically indicated.

“Blood pressure, lipid, and glycemic goals of therapy for people with diabetes apply equally to those who also have psychiatric disorders. However, all goals need to be individualized. The benefits and risks of different therapeutic agents used in the treatment of diabetes and its comorbidities should be considered in the context of the patient’s psychiatric condition and treatment.

“In summary, the panel recommends the following:

- Consideration of metabolic risks when starting SGAs
- Patient, family, and care giver education
- Baseline screening
- Regular monitoring
- Referral to specialized services, when appropriate”

**[Response Ends]**

**1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.**

**[Response Begins]**

**2022 and 2016 Submissions**

The guideline did not assign a grade to the quality of the quoted evidence.

**[Response Ends]**

**1a.06. Provide all other grades and definitions from the evidence grading system.**

**[Response Begins]**

**2022 and 2016 Submissions**

None.

**[Response Ends]**

**1a.07. Provide the grade assigned to the recommendation, with definition of the grade.**

**[Response Begins]**

**2022 and 2016 Submissions**

The guideline did not provide a grade for the cited recommendations.

**[Response Ends]**

**1a.08. Provide all other grades and definitions from the recommendation grading system.**

**[Response Begins]**

**2022 and 2016 Submissions**

None.

**[Response Ends]**

**1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.**

**[Response Begins]**

**2022 and 2016 Submissions**

This guideline was developed as a result of a consensus development conference of key experts and stakeholders. The key goal of the conference was to establish consensus on the following questions:

1. What is the current use of antipsychotic drugs?
2. What is the prevalence of obesity, pre-diabetes, and type 2 diabetes in the populations in which second-generation antipsychotics (SGAs) are used?
3. What is the relationship between the use of SGAs and the incidence of obesity or diabetes?
4. Given the above risks, how should patients be monitored for the development of significant weight gain, dyslipidemia, and diabetes, and how should they be treated if diabetes develops?
5. What research is needed to better understand the relationship between these drugs and significant weight gain, dyslipidemia, and diabetes?

While this is a consensus-based guideline, the authors cited 4 clinical practice guidelines, 5 systematic evidence reviews, 11 retrospective analyses, 3 randomized trials, and 1 cross-sectional analysis in support of the document. Evidence cited in support of the consensus document ranges from 1997 to 2003.

**[Response Ends]**

**1a.10. Provide the estimates of benefit, and consistency across studies.**

**[Response Begins]**

**2022 and 2016 Submissions**

The guideline was developed using a consensus-based approach involving a group of stakeholders and experts in the field. While the consensus-based approach was supplemented with a review of the evidence as part of the development of each guideline, details of the consistency of the reviewed evidence were not provided. This guideline does not provide a quantitative estimate of benefit for follow-up care for patients prescribed antipsychotic medications. However, there is consensus among the guidelines that close follow-up monitoring is an essential standard of care for patients prescribed antipsychotic medications to ensure effectiveness of treatment and to mitigate any adverse consequences or reactions to the drugs.

**[Response Ends]**

**1a.11. Indicate what, if any, harms were identified in the study.**

**[Response Begins]**

**2022 and 2016 Submissions**

While the guideline does not give a formal description or estimate of harms, we anticipate the expected benefits of follow-up care to far outweigh any potential harms because ongoing monitoring and follow-up is a basic standard of care for patients taking antipsychotic medications.

**[Response Ends]**

**1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.**

**[Response Begins]**

**2022 and 2016 Submissions**

We did not identify any new studies since the clinical guideline was published that change the conclusion in the recommendation.

**[Response Ends]**

Group 2 - Evidence - Systematic Reviews Table

**1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.**

**[Response Begins]**

**2022 Submission**

Keepers, G. A., Fochtman, L. J., Anzia, J. M., Benjamin, S., Lyness, J. M., Mojtabai, R., Servis, M., Walaszek, A., Buckley, P., Lenzenweger, M. F., Young, A. S., Degenhardt, A., Hong, S.-H., & (Systematic Review). (2020). The American Psychiatric Association Practice Guideline for the Treatment of Patients With Schizophrenia. *The American Journal of Psychiatry*, 177(9), 868–872. <https://doi.org/10.1176/appi.ajp.2020.177901>

**2016 Submission**

Lehman AF, Lieberman JA, Dixon LB, et al; American Psychiatric Association; Work Group on Schizophrenia. "Practice guideline for the treatment of patients with schizophrenia, second edition." *Am J Psychiatry*. 2004;161(2 Suppl):1-56. Available at [https://psychiatryonline.org/pb/assets/raw/sitewide/practice\\_guidelines/guidelines/schizophrenia.pdf](https://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/schizophrenia.pdf)

**[Response Ends]**

**1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.**

**[Response Begins]**

**2022 Submission**

"Determining the optimal dose of antipsychotic medication during acute treatment is complicated by the fact that there is usually a delay between initiation of treatment and full therapeutic response. Patients may take between 2 and 4

weeks to show an initial response and longer periods of time to show full or optimal response. Once a therapeutic dose of the antipsychotic medication is reached, overly rapid or premature escalation of medication doses can affect tolerability. Premature dose increases can also create the false impression of enhanced efficacy due to a higher dose when the observed response is actually related to elapsed time at a steady state level of medication. Available evidence suggests that patients who have not exhibited at least a 20% reduction in symptoms (or minimal improvement) by about 2 weeks on a therapeutic dose are unlikely to be much improved at 4–6 weeks as reflected by at least a 50% reduction in symptoms (Samara et al. 2015). Consequently, monitoring of the patient’s clinical status for 2–4 weeks is warranted on a therapeutic dose unless the patient is having uncomfortable side effects.”

**2016 Submission**

“The recommended dose is that which is both effective and not likely to cause side effects that are subjectively difficult to tolerate, since the experience of unpleasant side effects may affect long-term adherence [I]. The dose may be titrated as quickly as tolerated to the target therapeutic dose of the antipsychotic medication, and unless there is evidence that the patient is having uncomfortable side effects, monitoring of the patient’s clinical status for 2–4 weeks is warranted to evaluate the patient’s response to the treatment [II].”

**[Response Ends]**

**1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.**

**[Response Begins]**

**2022 Submission**

**High (denoted by the letter A)**=High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect.

**2016 Submission**

The guideline did not assign a grade to the quality of the quoted evidence.

**[Response Ends]**

**1a.06. Provide all other grades and definitions from the evidence grading system.**

**[Response Begins]**

**2022 Submission**

**Moderate (denoted by the letter B)**=Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate.

**Low (denoted by the letter C)**=Low confidence that the evidence reflects the true effect. Further research is likely to change our confidence in the estimate of effect and is likely to change the estimate.

**Insufficient**=Evidence that is unavailable or does not permit estimation of an effect.

**2016 Submission**

None.

**[Response Ends]**

**1a.07. Provide the grade assigned to the recommendation, with definition of the grade.**

**[Response Begins]**

**2022 Submission**

The recommendation has a grade of 1.

**2016 Submission**

The guideline assigned a grade II to its follow-up recommendation.

**[Response Ends]**

**1a.08. Provide all other grades and definitions from the recommendation grading system.**



**[Response Begins]**

**2022 Submission**

The APA grading scale is defined as follows:

“A recommendation (denoted by the numeral 1 after the guideline statement) indicates confidence that the benefits of the intervention clearly outweigh harms. A suggestion (denoted by the numeral 2 after the guideline statement) indicates greater uncertainty. Although the benefits of the statement are still viewed as outweighing the harms, the balance of benefits and harms is more difficult to judge, or the benefits or the harms may be less clear. With a suggestion, patient values and preferences may be more variable, and this can influence the clinical decision that is ultimately made.

“These strengths of recommendation correspond to ratings of strong or weak (also termed conditional) as defined under the Grading of Recommendations Assessment, Development and Evaluation (GRADE) method for rating recommendations in clinical practice guidelines.”

**2016 Submission**

The APA grading scale is defined as follows:

[I] Recommended with substantial clinical confidence.

[II] Recommended with moderate clinical confidence.

[III] May be recommended on the basis of individual circumstances.

**[Response Ends]**

**1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.**

**[Response Begins]**

**2022 Submission**

The evidence reviewed to support this guideline included clinical trials and meta-analyses related to schizophrenia and schizoaffective disorder to reflect a synthesis of the current literature and clinical practice on the treatment of patients with schizophrenia.

Evidence cited in support of this guideline includes 167 double-blind randomized clinical trials, 465 randomized clinical trials, and 402 randomized placebo-controlled trials. Evidence reviewed in the development of this guideline spanned from 1960 to 2018.

**2016 Submission**

The evidence reviewed to support this guideline included clinical trials and meta-analyses related to schizophrenia and schizoaffective disorder to reflect a synthesis of the current literature and clinical practice on the treatment of patients with schizophrenia.

Evidence cited in support of this guideline includes 181 double-blind randomized clinical trials, 116 randomized clinical trials, 152 clinical trials, 133 cohort or longitudinal studies, 122 case-control studies, 71 reviews with secondary data analysis, 167 literature reviews, and 497 other types of studies. Evidence reviewed in the development of this guideline spanned from 1994–2002.

**[Response Ends]**

**1a.10. Provide the estimates of benefit, and consistency across studies.**

**[Response Begins]**

**2022 Submission**

The APA guideline was developed based on a comprehensive literature review. However, details of the consistency of the reviewed evidence were not provided.

The guideline does not provide a quantitative estimate of benefit for follow-up care for patients prescribed antipsychotic medications. However, there is consensus among the guidelines that close follow-up monitoring is an essential standard of care for patients prescribed antipsychotic medications to ensure effectiveness of treatment and to mitigate any adverse consequences or reactions to the drugs.

**2016 Submission**

The APA guideline was developed based on a comprehensive literature review. However, details of the consistency of the reviewed evidence were not provided.

The guideline does not provide a quantitative estimate of benefit for follow-up care for patients prescribed antipsychotic medications. However, there is consensus among the guidelines that close follow-up monitoring is an essential standard

of care for patients prescribed antipsychotic medications to ensure effectiveness of treatment and to mitigate any adverse consequences or reactions to the drugs.

**[Response Ends]**

**1a.11. Indicate what, if any, harms were identified in the study.**

**[Response Begins]**

**2022 Submission**

While the guideline does not give a formal description or estimate of harms, we anticipate the expected benefits of follow-up care to far outweigh any potential harms because ongoing monitoring and follow-up is a basic standard of care for patients taking antipsychotic medications.

**2016 Submission**

While the guideline does not give a formal description or estimate of harms, we anticipate the expected benefits of follow-up care to far outweigh any potential harms because ongoing monitoring and follow-up is a basic standard of care for patients taking antipsychotic medications.

**[Response Ends]**

**1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.**

**[Response Begins]**

**2022 Submission**

We did not identify any new studies since the clinical guideline was published that change the conclusion in the recommendation.

**2016 Submission**

We did not identify any new studies since the clinical guideline was published that change the conclusion in the recommendation.

**[Response Ends]**

Group 3 - Evidence - Systematic Reviews Table

**1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.**

**[Response Begins]**

**2022 and 2016 Submissions**

University of South Florida College of Behavioral and Community Sciences. "2015 Florida Best Practice Psychotherapeutic Medication Guidelines for Adults." The University of South Florida, Florida Medicaid Drug Therapy Management Program sponsored by the Florida Agency for Health Care Administration, December 2015. Available at:

[http://www.medicaidmentalhealth.org/assets/file/Guidelines/Web\\_2015-Psychotherapeutic%20Medication%20Guidelines%20for%20Adults\\_Final\\_Approved1.pdf](http://www.medicaidmentalhealth.org/assets/file/Guidelines/Web_2015-Psychotherapeutic%20Medication%20Guidelines%20for%20Adults_Final_Approved1.pdf)

**[Response Ends]**

**1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.**

**[Response Begins]**

**2022 and 2016 Submissions**

**Principles of Practice**

*Comprehensive Assessment*

- "Careful, differential diagnostic evaluation
- Risk for suicide and violence
- Co-occurring mental and medical disorders

- Substance abuse disorders, including tobacco use
- Potential bipolar disorder must be assessed in patients presenting with depression
- Serious mental health conditions are chronic in nature; therefore, a long-term management plan is essential
  - Use measurement-based care to measure symptoms, side effects, and adherence
  - Select maintenance medications that have a low relative risk of weight gain and metabolic syndrome
  - Monitoring of physical health parameters and medication side effects (See Program publication A Summary for Monitoring Physical Health and Side-Effects of Psychiatric Medications in the Severely Mentally Ill Population available at [www.medicaidmentalhealth.org](http://www.medicaidmentalhealth.org))
  - Integrate care of psychiatrists and primary care providers
  - Incorporate collaborative/shared treatment decision-making with patients and family/caregivers
  - Perform a psychosocial assessment
  - Assess social support system (housing, family, other caregivers)
  - Evaluate threats to continuity of care (access to medication, adherence, etc.)
  - Give patients tools/support for recovery and self-management.”

*Adjunctive Psychosocial Treatments (As Indicated)*

- “Individual and family psychoeducation
- Cognitive-behavioral therapy (CBT)
- Interpersonal psychotherapy (IPT)
- Interpersonal and social rhythm therapy (IPSRT)
- Family-focused therapy
- Group psychoeducation (especially for bipolar disorder)
- Social skills training (especially in schizophrenia)
- Cognitive remediation/rehabilitation (to improve attention, memory, and/or executive function)

**\*\*Note on pharmacogenomic testing:** Limited data exists examining whether patient care that integrates pharmacogenomic test information results in better or safer treatment.”

*Measurement-Based Care*

“Questionnaires and rating scales are useful tools for diagnostic assessment and evaluation of treatment outcomes, and such instruments can be helpful in providing supplemental information to clinical judgment. The integration of measurement scales into routine clinical practice is suggested for each of the conditions covered in this document. Clinicians should use rating scales to assess symptom severity during the initial evaluation/treatment, when medication changes are implemented, and/or when the patient reports a change in symptoms.

- Treatment targets need to be precisely defined.
- Effectiveness and safety/tolerability of the medication treatment must be systematically assessed by methodical use of appropriate rating scales and side-effect assessment protocols.
- Internet links to the following scales are available on the program website -[www.medicaidmentalhealth.org](http://www.medicaidmentalhealth.org)
  - Beck Depression Inventory (BDI)
  - Brief Psychiatric Rating Scale (BPRS)
  - Clinical Global Impression (CGI) Scale
  - Clinician-Rated Dimensions of Psychosis Symptom Severity (CRDPSS)
  - Hamilton Rating Scale for Depression (HAM-D)
  - Montgomery-Asberg Depression Rating Scale (MADRS)
  - Patient Health Questionnaire (PHQ-9)
  - Positive and Negative Syndrome Scale (PANSS)

- Quick Inventory of Depression Symptomatology (QIDS)
- Young Mania Rating Scale (YMRS)”

**[Response Ends]**

**1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.**

**[Response Begins]**

**2022 and 2016 Submissions**

The guideline did not assign a grade to the quality of the quoted evidence.

**[Response Ends]**

**1a.06. Provide all other grades and definitions from the evidence grading system.**

**[Response Begins]**

**2022 and 2016 Submissions**

None.

**[Response Ends]**

**1a.07. Provide the grade assigned to the recommendation, with definition of the grade.**

**[Response Begins]**

**2022 and 2016 Submissions**

The guideline did not provide a grade for the cited recommendations.

**[Response Ends]**

**1a.08. Provide all other grades and definitions from the recommendation grading system.**

**[Response Begins]**

**2022 and 2016 Submissions**

None.

**[Response Ends]**

**1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.**

**[Response Begins]**

**2022 and 2016 Submissions**

The guideline was developed using a consensus-based approach. While evidence was reviewed as part of the development process for these guidelines, the details of the evidence review were not provided.

**[Response Ends]**

**1a.10. Provide the estimates of benefit, and consistency across studies.**

**[Response Begins]**

**2022 and 2016 Submissions**

The guideline was developed using a consensus-based approach involving a group of stakeholders and experts in the field. While the consensus-based approach was supplemented with a review of the evidence as part of the development of each guideline, details of the consistency of the reviewed evidence were not provided. The guideline does not provide a quantitative estimate of benefit for follow-up care for patients prescribed antipsychotic medications. However, there is consensus among the guidelines that close follow-up monitoring is an essential standard of care for patients prescribed antipsychotic medications to ensure effectiveness of treatment and to mitigate any adverse consequences or reactions to the drugs.

**[Response Ends]**

**1a.11. Indicate what, if any, harms were identified in the study.**

**[Response Begins]**

**2022 and 2016 Submissions**

While the guideline does not give a formal description or estimate of harms, we anticipate the expected benefits of follow-up care to far outweigh any potential harms because ongoing monitoring and follow-up is a basic standard of care for patients taking antipsychotic medications.

**[Response Ends]**

**1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.**

**[Response Begins]**

**2022 and 2016 Submissions**

We did not identify any new studies since the clinical guideline was published that change the conclusion in the recommendation.

**[Response Ends]**

Group 4 - Evidence - Systematic Reviews Table

**1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.**

**[Response Begins]**

**2022 and 2016 Submissions**

University of South Florida College of Behavioral and Community Sciences. "A Summary for Monitoring Physical Health and Side-Effects of Psychiatric Medications in the Severely Mentally Ill Population." The University of South Florida, Florida Medicaid Drug Therapy Management Program for Behavioral Health sponsored by the Florida Agency for Health Care Administration, March 2014. Available at:

[http://medicaidmentalhealth.org/\\_assets/file/Summaries/2014\\_Monitoring%20Physical%20Health%20and%20Side-Effects%20of%20Psychiatric%20Medicati....pdf](http://medicaidmentalhealth.org/_assets/file/Summaries/2014_Monitoring%20Physical%20Health%20and%20Side-Effects%20of%20Psychiatric%20Medicati....pdf)

**[Response Ends]**

**1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.**

**[Response Begins]**

**2022 and 2016 Submissions**

Assessment	Baseline	During titration/At target dose	Each Visit	At 6 Weeks	At 3 months	Every 3 months	At 12 months	Annually after first 12 months
Personal and family history	X	*	*	*	*	*	X	X

Assessment	Baseline	During titration/At target dose	Each Visit	At 6 Weeks	At 3 months	Every 3 months	At 12 months	Annually after first 12 months
Lifestyle behaviors	X	*	X	X	X	*	X	X
Weight	X	*	X	X	X	*	X	X
Waist Circumference	X	*	*	X	X	*	X	X
BP and pulse	X	X (during titration with clozapine and quetlapine)	X	X	X	*	X	X
Sedation/somnolence	X	*	X	X	X	*	X	X
Sexual/Reproductive dysfunction	X	X	*	X	X	X	*	*
Prolactin	X <sup>a</sup>	*	*	*	X <sup>b</sup>	*	X <sup>b</sup>	X <sup>b</sup>
Fasting blood glucose	X	*	*	X	X	*	X	X
Fasting lipid glucose	X	*	*	*	X	*	X	X
Parkinsonism (SAS or ESRS), Akathisia (AIMS or ESRS)	X	X	*	*	X	*	*	X
Electrolytes, full blood count, renal function	X	*	*	*	*	*	X (more frequent blood counts if on clozapine)	X (more frequent blood counts if on clozapine)
FTardive dyskensia	X	*	*	*	*	*	X	X
Liver function tests	X	*	*	*	*	*	X	X
Dental Health	X	*	*	*	*	*	X	X
ECG† parameters	X <sup>†</sup>	*	*	*	*	*	*	*

\* Cell intentionally left empty

<sup>1</sup> Adapted from Hert, et al, 2011. "Physical Illness in patients with severe mental disorders, II, Barriers to care, monitoring, and treatment guidelines, plus recommendations at the system and individual level. World Psychiatry. 10:138–151.

<sup>2</sup> Adapted from Florida Medicaid Drug Therapy Management Program for Behavioral Health: Florida Best Practice Medication Child

† Abbreviations: SAS = Simpson-Angus Scale; ESRS = Extrapyramidal Symptom Rating Scale; AIMS = Abnormal Involuntary Movement Scale.

‡ ECG = electrocardiogram; perform EKG at baseline then only if symptomatic.

#### [Response Ends]

**1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.**

#### [Response Begins]

##### 2022 and 2016 Submissions

The guideline did not assign a grade to the quality of the quoted evidence.

#### [Response Ends]

**1a.06. Provide all other grades and definitions from the evidence grading system.**

#### [Response Begins]

**2022 and 2016 Submissions**

None.

**[Response Ends]**

**1a.07. Provide the grade assigned to the recommendation, with definition of the grade.**

**[Response Begins]**

**2022 and 2016 Submissions**

The guideline did not provide a grade for the cited recommendations.

**[Response Ends]**

**1a.08. Provide all other grades and definitions from the recommendation grading system.**

**[Response Begins]**

**2022 and 2016 Submissions**

None.

**[Response Ends]**

**1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.**

**[Response Begins]**

**2022 and 2016 Submissions**

The guideline was developed using a consensus-based approach. While evidence was reviewed as part of the development process for these guidelines, the details of the evidence review were not provided.

**[Response Ends]**

**1a.10. Provide the estimates of benefit, and consistency across studies.**

**[Response Begins]**

**2022 and 2016 Submissions**

The guideline was developed using a consensus-based approach involving a group of stakeholders and experts in the field. While the consensus-based approach was supplemented with a review of the evidence as part of the development of each guideline, the consistency of the reviewed evidence were not provided.

None of the cited guidelines provide a quantitative estimate of benefit for follow-up care for patients prescribed antipsychotic medications. However, there is consensus among the guidelines that close follow-up monitoring is an essential standard of care for patients prescribed antipsychotic medications to ensure effectiveness of treatment and to mitigate any adverse consequences or reactions to the drugs.

**[Response Ends]**

**1a.11. Indicate what, if any, harms were identified in the study.**

**[Response Begins]**

**2022 and 2016 Submissions**

While the guideline does not give a formal description or estimate of harms, we anticipate the expected benefits of follow-up care to far outweigh any potential harms because ongoing monitoring and follow-up is a basic standard of care for patients taking antipsychotic medications.

**[Response Ends]**

**1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.**

**[Response Begins]**

**2022 and 2016 Submissions**

We did not identify any new studies since the clinical guideline was published that change the conclusion in the recommendation.

**[Response Ends]**

**1a.13. If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.**

**[Response Begins]**

Not applicable.

**[Response Ends]**

**1a.14. Briefly synthesize the evidence that supports the measure.**

**[Response Begins]**

**2022 Submission**

The literature also supports the underlying concept that follow-up visit with a provider is essential to monitor treatment effectiveness, evaluate health concerns, and adjust treatment as needed to minimize potential harms associated with the use of psychotropic medications. One 2014 cross-sectional analysis of nationally-representative data estimates that 35 percent to 50 percent of mental health care episodes consist of psychotropic drug fills without an outpatient visit to monitor treatment and up to 35 percent of episodes consisted of only a single visit (Le Cook et al. 2014).

Despite the importance of follow-up for patients taking antipsychotics, there is evidence that these patients are not receiving adequate follow-up care. For example, there is a growing body of evidence that shows persistent gaps in monitoring for metabolic effects of antipsychotic medications despite available guidelines and recommendations. While follow-up care should encompass more than just metabolic monitoring, metabolic testing rates can be useful to gain a general idea of the adequacy of follow-up care. In a 2016 analysis of data from the Missouri Medicaid program, Morrato and colleagues reported annual testing rates of 79.6 percent for glucose and 41.2 percent for lipids among beneficiaries taking antipsychotics (Morrato et al. 2016). This shows improvement over an earlier 2010 analysis data from three state Medicaid programs, which found testing rates as low as 27 percent for glucose testing and 10 percent for lipid testing (Morrato et al. 2010). This improvement is consistent with a 2011 analysis of Kansas Medicaid data that found improvement in annual testing between 2002 and 2007 from 23 percent to 75.3 percent for glucose monitoring and from 10.1 percent to 52.5 percent for lipid monitoring (Moeller, Rigler, Mayorga, Nazir, & Shireman 2011).

While progress on testing at a state level is encouraging, there is still considerable room for improvement at a local level. In a 2011 analysis of Medicaid data, rates of metabolic testing were found to vary significantly based on geographic location and patient characteristics such as age and comorbidity (Morrato et al. 2011). Inadequate follow-up care is often reflected by poor treatment adherence. A 2015 study found irregular attendance at follow-up appointments to be significantly associated with medication nonadherence (OR: 5.7; 95 percent confidence interval 2.92–11.31) among patients with psychiatric illness (Mert et al. 2015). A 2013 study found an antipsychotic non-adherence rate of nearly 38 percent among Medicaid patients with schizophrenia, with new prescription of antipsychotics and baseline non-adherence increasing the likelihood of non-adherence twelvefold (Lang et al. 2013). Appropriate follow-up care is essential for patients taking antipsychotic medications to receive the full benefit of treatment and to minimize potential harms associated with use of antipsychotics.

The importance of ongoing follow-up for patients on psychotropic medications is also emphasized in recent government efforts to promote best prescribing practices for psychotropic medications (MACPAC 2015). In a 2015 study, Mert and colleagues identified irregular follow-up as an important risk factor for medication non-adherence among patients with mental illness (Mert et al. 2015).

**2016 Submission**

In addition to the clinical guidelines reviewed above, the project team reviewed literature and consulted with stakeholders and clinical experts in development of this measure (see below). To define the measure specifications, the project team convened and consulted with a clinical advisory workgroup, including the follow-up period, target medications, and types of follow-up visits. The workgroup noted that timely follow-up is a minimal clinical standard of



care for patients with mental illness who are prescribed antipsychotics and other psychotropic medications and is a critical component of disease management.

The literature also supports the underlying concept that follow-up visit with a provider is essential to monitor treatment effectiveness, evaluate health concerns, and adjust treatment as needed to minimize potential harms associated with the use of psychotropic medications. One 2014 cross-sectional analysis of nationally-representative data estimates that 35 percent to 50 percent of mental health care episodes consist of psychotropic drug fills without an outpatient visit to monitor treatment and up to 35 percent of episodes consisted of only a single visit (Le Cook et al. 2014).

Despite the importance of follow-up for patients taking antipsychotics, there is evidence that these patients are not receiving adequate follow-up care. For example, there is a growing body of evidence that shows persistent gaps in monitoring for metabolic effects of antipsychotic medications despite available guidelines and recommendations. While follow-up care should encompass more than just metabolic monitoring, metabolic testing rates can be useful to gain a general idea of the adequacy of follow-up care. In a 2016 analysis of data from the Missouri Medicaid program, Morrato and colleagues reported annual testing rates of 79.6 percent for glucose and 41.2 percent for lipids among beneficiaries taking antipsychotics (Morrato et al. 2016). This shows improvement over an earlier 2010 analysis data from three state Medicaid programs, which found testing rates as low as 27 percent for glucose testing and 10 percent for lipid testing (Morrato et al. 2010). This improvement is consistent with a 2011 analysis of Kansas Medicaid data that found improvement in annual testing between 2002 and 2007 from 23 percent to 75.3 percent for glucose monitoring and from 10.1 percent to 52.5 percent for lipid monitoring (Moeller, Rigler, Mayorga, Nazir, & Shireman 2011).

While progress on testing at a state level is encouraging, there is still considerable room for improvement at a local level. In a 2011 analysis of Medicaid data, rates of metabolic testing were found to vary significantly based on geographic location and patient characteristics such as age and comorbidity (Morrato et al. 2011). Inadequate follow-up care is often reflected by poor treatment adherence. A 2015 study found irregular attendance at follow-up appointments to be significantly associated with medication nonadherence (OR: 5.7; 95 percent confidence interval 2.92-11.31) among patients with psychiatric illness (Mert et al. 2015). A 2013 study found an antipsychotic non-adherence rate of nearly 38 percent among Medicaid patients with schizophrenia, with new prescription of antipsychotics and baseline non-adherence increasing the likelihood of non-adherence twelvefold (Lang et al. 2013). Appropriate follow-up care is essential for patients taking antipsychotic medications to receive the full benefit of treatment and to minimize potential harms associated with use of antipsychotics.

The importance of ongoing follow-up for patients on psychotropic medications is also emphasized in recent government efforts to promote best prescribing practices for psychotropic medications (MACPAC 2015). In a 2015 study, Mert and colleagues identified irregular follow-up as an important risk factor for medication non-adherence among patients with mental illness (Mert et al. 2015).

#### **[Response Ends]**

#### **1a.15. Detail the process used to identify the evidence.**

##### **[Response Begins]**

##### **2022 Submission**

The project team conducted an environmental scan, which included a targeted literature review, an evaluation of existing performance measures related to physical and mental health care integration to identify critical measurement gaps, and interviews with key stakeholders and subject matter experts. Stakeholders interviewed by the project team emphasized the importance of ongoing follow-up after the prescription of psychotropic medications to evaluate treatment effectiveness and modify the treatment regimen as appropriate. Timely follow-up is also essential to address medication side effects and potential barriers to treatment adherence. As noted above, the project team consulted with a clinical advisory work group to identify the appropriate follow-up period, types of follow-up visits, and types of medications for inclusion in the measure.

##### **2016 Submission**

The project team conducted an environmental scan, which included a targeted literature review, an evaluation of existing performance measures related to physical and mental health care integration to identify critical measurement gaps, and interviews with key stakeholders and subject matter experts. Stakeholders interviewed by the project team emphasized the importance of ongoing follow-up after the prescription of psychotropic medications to evaluate treatment effectiveness and modify the treatment regimen as appropriate. Timely follow-up is also essential to address medication side effects and potential barriers to treatment adherence. As noted above, the project team consulted with a clinical advisory work group to identify the appropriate follow-up period, types of follow-up visits, and types of medications for inclusion in the measure.

**[Response Ends]**

**1a.16. Provide the citation(s) for the evidence.**

**[Response Begins]**

**2022 Submission**

De Hert M, Correll CU, Bobes J, et al. Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World Psychiatry*.2011;10:52-77.

Lang K, Federico V, Muser E, Menzin J, Menzin J. Rates and predictors of antipsychotic non-adherence and hospitalization in Medicaid and commercially-insured patients with schizophrenia. *J Med Econ*. 2013;16(8):997-1006. doi: 10.3111/13696998.2013.816310.

Le Cook B, Zuvekas SH, Carson N, et al. Assessing racial/ethnic disparities in treatment across episodes of mental health care. *Health Serv Res*.2014;49(1):206-29. doi: 10.1111/1475-6773.12095.

Medicaid and CHIP Payment and Access Commission (MACPAC). Report to Congress on Medicaid and CHIP. June 2015. available at: <https://www.macpac.gov/wp-content/uploads/2015/06/June-2015-Report-to-Congress-on-Medicaid-and-CHIP.pdf>. Accessed February 4, 2016.

Mert DG, Turgut NH, Lelleci M, Semiz M. Perspectives on reasons of medication nonadherence in psychiatric patients. *Patient Prefer Adherence*. 2015;9:87-93. doi: 10.2147/PPA.S75013.

Moeller KE, Rigler SK, Mayorga A, Nazir N, and Shireman TI. Quality of monitoring for metabolic effects associated with second generation antipsychotics in patients with schizophrenia on public insurance. *Schizophr Res*. 2011;126(1-3):117-23. doi: 10.1016/j.schres.2010.11.015

Morrato EH, Campagna EJ, Brewer SE, et al. Metabolic testing for adults in a state Medicaid program receiving antipsychotics: remaining barriers to achieving population health prevention goals. *JAMA Psychiatry*.2016;73(7):721-30. doi:10.1001/jamapsychiatry.2016.0538

Morrato EH, Druss B, Hartung DM, et al. Metabolic testing rates in 3 state Medicaid programs after FDA warnings and ADA/APA recommendations for second-generation antipsychotic drugs. *Arch Gen Psychiatry*. 2010;67(1):17-24. doi:10.1001/archgenpsychiatry.2009.179

Morrato EH, Druss BG, Hartung DM, et al. Small area variation and geographic and patient-specific determinants of metabolic testing in antipsychotic users. *Pharmacoepidemiol Drug Saf*. 2011;20(1):66-75. doi:10.1002/pds.2062.

**2016 Submission**

De Hert M, Correll CU, Bobes J, et al. Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. *World Psychiatry*.2011;10:52-77.

Lang K, Federico V, Muser E, Menzin J, Menzin J. Rates and predictors of antipsychotic non-adherence and hospitalization in Medicaid and commercially-insured patients with schizophrenia. *J Med Econ*. 2013;16(8):997-1006. doi: 10.3111/13696998.2013.816310.

Le Cook B, Zuvekas SH, Carson N, et al. Assessing racial/ethnic disparities in treatment across episodes of mental health care. *Health Serv Res*.2014;49(1):206-29. doi: 10.1111/1475-6773.12095.

Medicaid and CHIP Payment and Access Commission (MACPAC). Report to Congress on Medicaid and CHIP. June 2015. available at: <https://www.macpac.gov/wp-content/uploads/2015/06/June-2015-Report-to-Congress-on-Medicaid-and-CHIP.pdf>. Accessed February 4, 2016.

Mert DG, Turgut NH, Lelleci M, Semiz M. Perspectives on reasons of medication nonadherence in psychiatric patients. *Patient Prefer Adherence*. 2015;9:87-93. doi: 10.2147/PPA.S75013.

Moeller KE, Rigler SK, Mayorga A, Nazir N, and Shireman TI. Quality of monitoring for metabolic effects associated with second generation antipsychotics in patients with schizophrenia on public insurance. *Schizophr Res*. 2011;126(1-3):117-23. doi: 10.1016/j.schres.2010.11.015

Morrato EH, Campagna EJ, Brewer SE, et al. Metabolic testing for adults in a state Medicaid program receiving antipsychotics: remaining barriers to achieving population health prevention goals. *JAMA Psychiatry*.2016;73(7):721-30. doi:10.1001/jamapsychiatry.2016.0538

Morrato EH, Druss B, Hartung DM, et al. Metabolic testing rates in 3 state Medicaid programs after FDA warnings and ADA/APA recommendations for second-generation antipsychotic drugs. *Arch Gen Psychiatry*. 2010;67(1):17-24. doi:10.1001/archgenpsychiatry.2009.179

Morrato EH, Druss BG, Hartung DM, et al. Small area variation and geographic and patient-specific determinants of metabolic testing in antipsychotic users. *Pharmacoepidemiol Drug Saf*. 2011;20(1):66-75. doi:10.1002/pds.2062.

**[Response Ends]**

**1b.01. Briefly explain the rationale for this measure.**

*Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.*

**[Response Begins]**

Among individuals with serious mental illness, physical health problems such as cardiovascular disease, metabolic disorders, and infectious disease are more prevalent compared to the general population. Antipsychotic medications can exacerbate existing physical problems as well as increase a patient's risk for developing new health concerns such as metabolic complications. Timely follow-up with a provider following the prescription of antipsychotic medications is an essential first step to ensure that physical impacts of antipsychotic medications are identified and addressed early. Early follow up is also critical to monitor for treatment effectiveness and modify dosage as necessary, as well as to identify and address any barriers to treatment adherence. By proactively following up with patients who are prescribed antipsychotic medications, providers can identify problems early in the course of treatment and minimize potential harms associated with use of those medications. Regardless of the care setting in which a patient is being treated, comprehensive assessment of both physical and mental health factors is an essential aspect of treatment with antipsychotic medications.

**[Response Ends]**

**1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.**

*Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

**[Response Begins]**

The measure was tested using the Transformed Medicaid Statistical Information System (T-MSIS) data and Medicare Part A, B, C and D claims data from January 1, 2018 through December 31, 2018. Thirteen states representing a cross section of US regions (2 Northeast, 1 Midwest, 4 South, 6 West), were used for measure testing: State 1 W, State 2 S, State 3 S, State 4 W, State 5 MW, State 6 W, State 7 NE, State 8 W, State 9 W, State 10 S, State 11 S, State 12 NE, and State 13 W. These states have been blinded to protect confidentiality.

For the 13 states used in measure testing, there were 285,060 new antipsychotic prescriptions for 273,078 beneficiaries. Beneficiaries who received an outpatient follow-up visit within 28-days of their new antipsychotic prescription fill date are compliant for this measure.

The number of beneficiaries, number of new antipsychotic prescriptions (denominator), and the measure performance rate overall and for each state are shown below. State-level measure performance rates ranged from 35.86% to 58.72%

**Overall (13 states):**

Number of eligible beneficiaries: 273,078

Number of new antipsychotic prescriptions (denominator): 285,060

Weighted\* mean performance rate: 45.82%

Unweighted mean performance rate: 46.72%

Standard deviation: 6.04%

Minimum: 35.86%

25th Percentile: 43.62%

50th Percentile: 45.17%

75th Percentile: 50.68%

Maximum: 58.72%

Interquartile Range (IQR): 7.10%

\*Weighted average across states by denominator (number of new antipsychotic prescriptions)

**State-level results:**

*State 1 West*

- Number of eligible beneficiaries: 32,065
- Number of new antipsychotic prescriptions (denominator): 33,690
- Performance Rate: 53.6%
- 95% CI: 53.1%, 54.1%

*State 2 South*

- Number of eligible beneficiaries: 6,277

- Number of new antipsychotic prescriptions (denominator): 6,771
- Numerator: 2,428
- Performance Rate: 35.9%
- 95% CI: 34.7%, 37.0%

*State 3 South*

- Number of eligible beneficiaries: 100,884
- Number of new antipsychotic prescriptions (denominator): 104,074
- Numerator: 46,260
- Performance Rate: 44.5%
- 95% CI: 44.2%, 44.8%

*State 4 West*

- Number of eligible beneficiaries: 4,781
- Number of new antipsychotic prescriptions (denominator): 5,080
- Numerator: 2,426
- Performance Rate: 47.8%
- 95% CI: 46.4%, 49.1%

*State 5 Midwest*

- Number of eligible beneficiaries: 23,349
- Number of new antipsychotic prescriptions (denominator): 24,478
- Numerator: 10,991
- Performance Rate: 44.9%
- 95% CI: 44.3%, 45.5%

*State 6 West*

- Number of eligible beneficiaries: 5,067
- Number of new antipsychotic prescriptions (denominator): 5,310
- Numerator: 2,273
- Performance Rate: 42.8%
- 95% CI: 41.5%, 44.1%

*State 7 Northeast*

- Number of eligible beneficiaries: 3,371
- Number of new antipsychotic prescriptions (denominator): 3,509
- Numerator: 1,647
- Performance Rate: 46.9%
- 95% CI: 45.3%, 48.6%

*State 8 West*

- Number of eligible beneficiaries: 11,953
- Number of new antipsychotic prescriptions (denominator): 12,619
- Numerator: 5,505
- Performance Rate: 43.6%
- 95% CI: 42.8%, 44.5%

*State 9 West*

- Number of eligible beneficiaries: 11,581
- Number of new antipsychotic prescriptions (denominator): 12,142
- Numerator: 6,384
- Performance Rate: 52.6%
- 95% CI: 51.7%, 53.5%

*State 10 South*

- Number of eligible beneficiaries: 12,801
- Number of new antipsychotic prescriptions (denominator): 13,569
- Numerator: 6,877
- Performance Rate: 50.7%
- 95% CI: 49.8%, 51.5%

*State 11 South*

- Number of eligible beneficiaries: 28,877
- Number of new antipsychotic prescriptions (denominator): 29,985

- Numerator: 12,064
- Performance Rate: 40.2%
- 95% CI: 39.7%, 40.8%

*State 12 Northeast*

- Number of eligible beneficiaries: 2,980
- Number of new antipsychotic prescriptions (denominator): 3,176
- Numerator: 1,865
- Performance Rate: 58.7%
- 95% CI: 57.0%, 60.4%

*State 13 West*

- Number of eligible beneficiaries: 29,092
- Number of new antipsychotic prescriptions (denominator): 30,657
- Numerator: 13,849
- Performance Rate: 45.2%
- 95% CI: 44.6%, 45.7%

**[Response Ends]**

**1b.03. If no or limited performance data on the measure as specified is reported above, 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. Include citations.**

**[Response Begins]**

Not applicable. Data have been included for Question 1b.02.

**[Response Ends]**

**1b.04. 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.**

*Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., “topped out”, disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

**[Response Begins]**

During measure testing, the measure performance rates were stratified by age group, race/ethnicity, dual eligibility for Medicare and Medicaid, gender, and Medicaid eligibility category.

**Measure performance for the 13 states, stratified by age group:**

*Age 18–34 years*

- Denominator: 80,269
- Numerator: 31,173
- Performance rate: 38.8%

*Age 35–44 years*

- Denominator: 51,416
- Numerator: 23,372
- Performance rate: 45.5%

*Age 45–54 years*

- Denominator: 55,680
- Numerator: 27,726
- Performance rate: 49.8%

*Age 55–64 years*

- Denominator: 53,447
- Numerator: 27,829
- Performance rate: 52.1%

*Age 65–74 years*

- Denominator: 22,382
- Numerator: 11,570
- Performance rate: 51.7%

*Age 75 years and older*

- Denominator: 21,886
- Numerator: 8,951
- Performance rate: 40.9%

**Measure performance for the 13 states, stratified by race/ethnicity**

*White non-Hispanic*

- Denominator: 148,992
- Numerator: 71,322
- Performance rate: 47.9%

*Black Non-Hispanic*

- Denominator: 52,749
- Numerator: 21,300
- Performance rate: 40.4%

*Hispanic*

- Denominator: 44,453
- Numerator: 21,014
- Performance rate: 47.3%

*Other/Unknown Race and Ethnicity*

- Denominator: 38,936
- Numerator: 16,985
- Performance rate: 43.6%

**Measure performance for the 13 states, stratified by dual eligibility**

*Medicaid-Only: Not Dually Eligible For Medicare and Medicaid*

- Denominator: 179,323
- Numerator: 79,332
- Performance rate: 44.2%

*Dually Eligible for Medicare and Medicaid*

- Denominator: 105,737
- Numerator: 51,289
- Performance rate: 48.5%

**Measure performance for the 13 states, stratified by gender**

*Female*

- Denominator: 172,539
- Numerator: 84,279
- Performance rate: 48.9%

*Male*

- Denominator: 112,521
- Numerator: 46,342
- Performance rate: 41.2%

**Measure performance for the 13 states, stratified by Medicaid beneficiary category**

*Aged*

- Denominator: 43,132
- Numerator: 20,011
- Performance rate: 46.4%

*Disabled*

- Denominator: 146,598
- Numerator: 65,934
- Performance rate: 45.0%

*Adult*

- Denominator: 89,012
- Numerator: 42,242
- Performance rate: 47.5%

CHIP/Child/Unknown

- Denominator: 6,318
- Numerator: 2,434
- Performance rate: 38.5%

**Data Sources:** Transformed Medicaid Statistical Information System (T-MSIS) Analytic files (TAF) Research Identifiable Files (RIFs): demographic and eligibility (DE), other services (OT), inpatient (IP) and pharmacy (Rx). Medicare Advantage (MA) encounter files and Medicare Fee-For-Service files: master beneficiary summary file (MBSF), carrier, outpatient (OP), inpatient (IP), MedPAR and prescription drug event (PDE).

[Response Ends]

**1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.**

[Response Begins]

Not applicable. Performance data provided for Question **1b.4**.

[Response Ends]

## Criteria 2: Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.

---

**spma.01. Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.**

[Response Begins]

No

[Response Ends]

**spma.02. Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.**

**For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.**

*For example, specifications may have been updated based on suggestions from a previous NQF CDP review.*

[Response Begins]

Not applicable

[Response Ends]

**sp.01. Provide the measure title.**

*Measure titles should be concise yet convey who and what is being measured (see [What Good Looks Like](#)).*

[Response Begins]

Follow-Up Care for Adult Medicaid Beneficiaries Who are Newly Prescribed an Antipsychotic Medication

[Response Ends]

**sp.02. Provide a brief description of the measure.**

*Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).*

**[Response Begins]**

Percentage of new antipsychotic prescriptions for Medicaid beneficiaries age 18 years and older who have completed a follow-up visit with a provider with prescribing authority within four weeks (28 days) of prescription of an antipsychotic medication.

**[Response Ends]**

**sp.04. Check all the clinical condition/topic areas that apply to your measure, below.**

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Surgery: General*

**[Response Begins]**

Behavioral Health: Bipolar Disorder  
Behavioral Health: Other Serious Mental Illness  
Behavioral Health: Schizophrenia  
Other (specify)

**[Other (specify) Please Explain]**

Home and Community-Based Services, including older adults, persons with a physical disability, persons with an intellectual or developmental disability (ID/DD), persons with an acquired brain injury (ABI), and persons with mental health or substance use disorders (MH/SUD).

**[Response Ends]**

**sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.**

**[Response Begins]**

Care Coordination  
Safety: Medication

**[Response Ends]**

**sp.06. Select one or more target population categories.**

*Select only those target populations which can be stratified in the reporting of the measure's result.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Populations at Risk: Populations at Risk*

**[Response Begins]**

Adults (Age >= 18)  
Populations at Risk: Dual eligible beneficiaries of Medicare and Medicaid  
Populations at Risk: Populations at Risk

**[Response Ends]**

**sp.07. Select the levels of analysis that apply to your measure.**

*Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.*



Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- Clinician: Clinician
- Population: Population

**[Response Begins]**

Population: Regional and State

**[Response Ends]**

**sp.08. Indicate the care settings that apply to your measure.**

Check *ONLY* the settings for which the measure is SPECIFIED and TESTED.

**[Response Begins]**

Outpatient Services

**[Response Ends]**

**sp.09. 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. If no URL is available, indicate "none available".

**[Response Begins]**

<https://www.medicaid.gov/state-resource-center/innovation-accelerator-program/iap-downloads/functional-areas/techspecsmanual-nqf-3313.pdf>

**[Response Ends]**

**sp.11. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.**

Attach an excel or csv file; if this poses an issue, [contact staff](#). Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

**[Response Begins]**

Available in attached Excel or csv file

**[Response Ends]**

Attachment: 3313\_NQF-3313\_Value\_Sets.xlsx

**sp.12. State the numerator.**

Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).

DO NOT include the rationale for the measure.

**[Response Begins]**

The percentage of Medicaid beneficiaries aged 18 years and older with new antipsychotic prescriptions who completed an outpatient follow-up visit with a provider with prescribing authority within 28 days of the new antipsychotic prescription fill.

**[Response Ends]**

**sp.13. Provide details needed to calculate the numerator.**

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

The numerator uses a 28-day (4 week) follow-up period based on clinical guidelines for appropriate follow-up after prescription of new antipsychotic medications. The optimal follow-up period was determined through testing and consultation with the Clinical Advisory Work group. The day after the prescription is filled is counted as day 1 of the follow-up period. The date of the follow-up visit with a provider is determined by using the service date on the medical claim.

See attached Excel file for CPT and HCPCS codes that qualify for the numerator.

**[Response Ends]**

**sp.14. State the denominator.**

*Brief, narrative description of the target population being measured.*

**[Response Begins]**

New antipsychotic prescriptions for Medicaid beneficiaries aged 18 years and older.

**[Response Ends]**

**sp.15. Provide details needed to calculate the denominator.**

*All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.*

*Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

Target population meets the following conditions:

1. Medicaid beneficiary aged 18 years and older (including dual-eligible and Medicaid-only enrollees)
2. Newly prescribed an antipsychotic medication
3. Enrolled in Medicaid during the 120 days (four months) prior to and the 28 days (4 weeks) following a new prescription of an antipsychotic medication

Beneficiaries with “newly filled prescriptions” are those who have had no antipsychotic medications dispensed for either new or refill antipsychotic prescription during a period of 120 days (four months) prior to the prescription fill date. The measure focuses on new prescriptions of antipsychotic medications.

We used National Drug Codes to identify the following antipsychotic medications for this measure:

- Amitriptyline Hydrochloride
- Aripiprazole (Abilify)
- Aripiprazole lauroxil (Aristada)
- Asenapine Maleate (Saphris)
- Brexpiprazole (Rexulti)
- Cariprazine hydrochloride (Vraylar)
- Chlorpromazine hydrochloride
- Clozapine (Clozaril, Fazaclo, Versacloz)
- Droperidol (Inapsine)
- Fentanyl citrate/droperidol
- Fluoxetine Hydrochloride-Olanzapine (Symbyax, Fluoxetine, Sarafem, Selfaemra)
- Fluoxetine-olanzapine
- Fluphenazine
- Haloperidol (Haldol)
- Iloperidone (Fanapt)
- Loxapine succinate (Loxitane, Adasuve)

- Lurisdone hydrochloride (Latuda)
- Molindone hydrochloride (Moban)
- Olanzapine (Zyprexa)
- Paliperidone (Invega)
- Perphenazine
- Pimozide (Orap)
- Prochlorperazine maleate (Compazine, Compro)
- Quetiapine fumarate (Seroquel)
- Risperidone (Risperdal)
- Thioridazine hydrochloride
- Thiothixene (Navane)
- Trifluoperazine hydrochloride
- Ziprasidone (Geodon)

See attached Excel file for NDCs that qualify for the denominator.

**[Response Ends]**

**sp.16. Describe the denominator exclusions.**

*Brief narrative description of exclusions from the target population.*

**[Response Begins]**

- Medicaid beneficiaries with an acute inpatient admission during the 28-day follow-up period after prescription of an antipsychotic medication
- Patients who expired within 28 days of new antipsychotic prescription fill date.

**[Response Ends]**

**sp.17. Provide details needed to calculate the denominator exclusions.**

*All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.*

**[Response Begins]**

Acute inpatient admission during the 28-day follow-up period: Beneficiaries with an inpatient admission during the 28-day follow-up period are excluded from the measure.

Death: Patients with a date of death during the 28-day follow-up period are excluded from the measure.

**[Response Ends]**

**sp.18. Provide all information required to stratify the measure results, if necessary.**

*Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.*

**[Response Begins]**

Not applicable; this measure is not stratified.

**[Response Ends]**

**sp.19. Select the risk adjustment type.**

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

**[Response Begins]**

No risk adjustment or risk stratification

**[Response Ends]**

**sp.20. Select the most relevant type of score.**

*Attachment: If available, please provide a sample report.*

**[Response Begins]**

Rate/proportion

**[Response Ends]**

**sp.21. Select the appropriate interpretation of the measure score.**

*Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score*

**[Response Begins]**

Better quality = Higher score

**[Response Ends]**

**sp.22. Diagram or describe the calculation of the measure score as an ordered sequence of steps.**

*Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.*

**[Response Begins]**

Step 1: Identify Eligible Population

Step 1A. Identify Medicaid beneficiaries (both dual-eligible and Medicaid-only enrollees) aged 18 years and older.

Step 1B. From this group, identify new prescriptions of one or more antipsychotic medications.

Step 2: Apply Continuous Enrollment Requirement

From the population identified in step 1:

Step 2A. Remove any prescriptions for beneficiaries who were not continuously enrolled for at least 120 days before or 28 days following the new prescription.

Step 3: Apply Exclusions:

From the population identified in step 2:

Step 3A. Remove any prescriptions for beneficiaries who had an acute inpatient admission during the 28 days following the new prescription.

Step 3B. Remove any prescriptions for beneficiaries who expired during the 28 days following the new prescription.

Step 4: Numerator

From the prescriptions within the denominator (after denominator exclusions have been applied):

Step 4A. Identify the number of prescriptions for beneficiaries who had a qualifying outpatient encounter within 28 days of the prescription date of the antipsychotic medication.

Step 5: Calculate the measure score

Step 5A. Divide the total number of prescriptions in the numerator by the total number of prescriptions in the denominator, after denominator exclusions have been applied.

Step 5B. Multiply this number by 100 to determine the performance rate.

**[Response Ends]**

**sp.25. If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.**

**[Response Begins]**

Not applicable; this measure does not use a sample.

**[Response Ends]**

**sp.28. Select only the data sources for which the measure is specified.**

**[Response Begins]**

Claims

**[Response Ends]**

**sp.29. Identify the specific data source or data collection instrument.**

*For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.*

**[Response Begins]**

We used the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAFs) Research Identifiable Files (RIFs) which contains beneficiary, service utilization, administrative claims, and expenditure data for the Medicaid population, including those covered through both fee-for-service (FFS) and managed care payers. The Medicaid T-MSIS TAF RIFs from September 1, 2017 through December 31, 2018 were used to assess importance, reliability and validity: Demographic and Eligibility (DE), Other Services (OT), Inpatient (IP) and Pharmacy (RX). For beneficiaries dually enrolled in Medicare [Medicare Advantage (MA) or Medicare Fee-For-Service (FFS)], in addition to the T-MSIS TAF RIFs data, we also used Medicare Part A, B, C and D administrative claims (September 1, 2017 through December 31, 2018) from the Center for Medicare and Medicaid's Chronic Conditions Warehouse (CCW): MA encounter and Medicare FFS carrier, Outpatient (OP), Inpatient (IP), MedPAR and Prescription Drug Event (PDE) files.

Measure validity was assessed by conducting a convergent validity analysis using NCQA Healthcare Effectiveness Data and Information Set (HEDIS) data from measure year 2018. HEDIS data are collected from Medicaid Health Management Organizations and Preferred Provider Organizations via the NCQA Interactive Data Submission System (IDSS) portal.

**[Response Ends]**

**sp.30. Provide the data collection instrument.**

**[Response Begins]**

No data collection instrument provided

**[Response Ends]**

**2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).**

*Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:*

**Current Submission:**

*Updated testing information here.*

**Previous Submission:**

*Testing from the previous submission here.*

**[Response Begins]**

Yes

**[Response Ends]**

**2ma.02. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).**

**Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:**

**Current Submission:**

Updated testing information here.

**Previous Submission:**

Testing from the previous submission here.

**[Response Begins]**

Yes

**[Response Ends]**

**2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?**

**[Response Begins]**

No

**[Response Ends]**

**2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.**

**Please update the Scientific Acceptability: Validity - Other Threats to Validity section.**

**Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.**

**[Response Begins]**

No additional risk adjustment analysis included

**[Response Ends]**

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- 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.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the [Submitting Standards webpage](#).
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the [2021 Measure Evaluation Criteria and Guidance](#).

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

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

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

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

AND

If patient preference (e.g., informed decision-making) 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).

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

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

OR

- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

## Definitions

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).

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 measure scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

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.

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

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

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.

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Scientific Acceptability sections. For example:

**2021 Submission:**

Updated testing information here.

**2018 Submission:**

Testing from the previous submission here.

**2a.01. Select only the data sources for which the measure is tested.**

**[Response Begins]**

Claims

**[Response Ends]**

**2a.02. 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).*

**[Response Begins]**

**2021 Submission**

The Transformed Medicaid Statistical Information System (T-MSIS) contains beneficiary, service utilization, administrative claims, and expenditure data for the Medicaid population, including those covered through both fee-for-service (FFS) and managed care payers. The team used the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF) Research Identifiable Files (RIFs) which have beneficiary-level claims and enrollment data for the Medicaid population covered through fee-for-service (FFS) and managed care payers. We used T-MSIS data from September 1, 2017, through December 31, 2017, as the measure includes a 120-day look-back period for continuous enrollment and to identify new antipsychotic prescriptions, as well as T-MSIS data from calendar year 2018. Among the available T-MSIS files, the following were accessed to assess importance, reliability, and validity.

- Medicaid Demographic and Eligibility (DE) file;
- Medicaid Inpatient (IP) file;
- Medicaid Other Services (OT) file; and
- Medicaid Pharmacy (RX) file.

In addition to the T-MSIS TAF RIFs, for Medicaid beneficiaries dually eligible for Medicare and Medicaid, the following Medicare Advantage (MA) and Medicare FFS 2017 and 2018 RIFs were accessed to assess importance, reliability, and validity:

- Master Beneficiary Summary File (MBSF);
- Medicare Prescription Drug Event file;
- Medicare MedPAR file;
- Medicare Outpatient Encounter file;
- Medicare Carrier Encounter file;
- Medicare Inpatient Encounter file;
- Medicare Outpatient file; and
- Medicare Carrier file.

For validity analysis, the team calculated the correlation between NQF 3313 and related measures for the behavioral health population in the Healthcare Effectiveness Data and Information Set (HEDIS). HEDIS data are collected from Medicaid Health Management Organizations and Preferred Provider Organizations via the National Committee for Quality Assurance (NCQA) Interactive Data Submission System (IDSS) portal. The team used HEDIS data from measurement year 2018.

**2016 Submission**



For the beneficiaries dually eligible for both Medicare and Medicaid, we also used Medicare enrollment data from the Medicare Beneficiary Summary File (MBSF) and from the following sources of claims data in CMS’s Data Extract System (DESY):

- Part D Prescription Drug claims from CMS’s Prescription Drug Events file
- Carrier or Physician/Supplier Part B claims from the National Claims History (NCH) file for final action, fee-for-service (FFS) claims submitted on a CMS-1500 claim form from non-institutional providers such as physicians, physician assistants, clinical social workers, and nurse practitioners
- Outpatient Department (OPD) claims from the NCH for final action, FFS claims data submitted by institutional outpatient providers such as hospital outpatient departments, rural health clinics, renal dialysis facilities, outpatient rehabilitation facilities, comprehensive outpatient facilities, and community mental health centers
- Medicare Provider Analysis Review (MedPAR) files for final action acute hospital and skilled nursing facility (SNF) admissions

**[Response Ends]**

**2a.03. Provide the dates of the data used in testing.**

*Use the following format: “MM-DD-YYYY - MM-DD-YYYY”*

**[Response Begins]**

**2021 Submission**

09-01-2017–12-31-2018

We tested the measure (referred to here as “NQF 3313”) using Medicaid and Medicare data from calendar year 2017 and 2018 (the measurement year). The measure requires a 16-week look-back period to establish that a beneficiary did not have any other antipsychotic medications prescribed; thus, beneficiaries are required to be enrolled in Medicaid for the 16 weeks before the prescription was filled. To account for the look-back period and establish continuous enrollment during this period, we also used data from September 1, 2017, through December 31, 2017.

**2016 Submission**

We tested the measure (referred to here as “PMH-1NQF 3313”) using Medicaid and Medicare data from calendar year 2014 (the measurement year). The measure requires a 16-week look-back period to establish that a beneficiary did not have any other antipsychotic medications prescribed; thus, beneficiaries are required to be enrolled in Medicaid for the 16 weeks before the prescription was filled. To account for the look-back period and establish continuous enrollment during this period, we also used data from September 1, 2013, through December 31, 2013.

**[Response Ends]**

**2a.04. Select the levels of analysis for which the measure is tested.**

*Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.*

*Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.*

*Please do not select:*

- *Clinician: Clinician*
- *Population: Population*

**[Response Begins]**

Population: Regional and State

**[Response Ends]**

**2a.05. List the measured entities 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.*

**[Response Begins]**

**2021 Submission**

Thirteen states representing a cross section of US regions (2 Northeast, 1 Midwest, 4 South, 6 West), were used for measure testing: State 1 W, State 2 S, State 3 S, State 4 W, State 5 MW, State 6 W, State 7 NE, State 8 W, State 9 W, State 10 S, State 11 S, State 12 NE, and State 13 W. These states have been blinded to protect confidentiality.

**2016 Submission**

15 states: State A, State B, State C, State D, State E, State F, State 5, State G, State H, State I, State J, State K, State 12, State L, and State M.

**[Response Ends]**

**2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.**

*If there is a minimum case count used for testing, that minimum must be reflected in the specifications.*

**[Response Begins]**

**2021 Submission**

For Medicaid-only and Medicare-Medicaid dual eligible beneficiaries age 18 years and older in these thirteen states, there were 5,074,895 antipsychotic prescriptions filled during the measure period (January 1, 2018 through November 30, 2018). Of these, 362,341 met the definition of a “new” antipsychotic prescription, and 285,060 met the criteria for inclusion in the measure (beneficiaries continuously enrolled in Medicaid and did not die or have an inpatient hospitalization during the 28-day follow-up period). Measure performance, defined as the proportion of new antipsychotic prescriptions filled by Medicaid beneficiaries who received an outpatient follow-up visit within 28-days of their prescription fill date) was 45.8 percent (130,621 follow-ups to 285,060 new antipsychotic prescriptions). Table 1 describes the prescriptions included in the analytic sample.

Table 1. Analytic sample selection (1/1/2018–11/30/2018)

*	Number of Medicaid-only beneficiaries (N)	Number of dual-eligible beneficiaries (N)	Number of Medicaid-only and dual-eligible beneficiaries age 18+ years as of 1/1/18 (N)	Number of antipsychotic prescriptions 1/1/18 - 11/30/18 (N)	Number of Medicaid beneficiaries age 18+ years with an antipsychotic prescription 1/1/18 - 11/30/18 (N)	Number of new antipsychotic prescriptions 1/1/18 - 11/30/18 (N)	Number of new antipsychotic prescriptions that meet the criteria for inclusion in the measure (Denom) (N)
Level of Analysis	Beneficiary	Beneficiary	Beneficiary	Prescription	Beneficiary	Prescription	Prescription
<b>All 13 states</b>	<b>14,350,178</b>	<b>2,434,045</b>	<b>8,872,320</b>	<b>5,074,895</b>	<b>698,870</b>	<b>362,341</b>	<b>285,060</b>
State 1 W	2,134,551	261,853	1,434,200	634,430	84,685	42,193	33,690
State 2 S	254,184	39,900	201,624	94,397	14,593	8,001	6,771
State 3 S	4,137,101	969,364	2,485,875	1,302,820	203,301	122,796	104,074
State 4 W	377,601	48,421	260,568	99,052	13,155	6,205	5,080
State 5 MW	1,035,820	208,780	535,538	761,531	84,658	34,359	24,478

<b>*</b>	<b>Number of Medicaid-only beneficiaries (N)</b>	<b>Number of dual-eligible beneficiaries (N)</b>	<b>Number of Medicaid-only and dual-eligible beneficiaries age 18+ years as of 1/1/18 (N)</b>	<b>Number of antipsychotic prescriptions 1/1/18 - 11/30/18 (N)</b>	<b>Number of Medicaid beneficiaries age 18+ years with an antipsychotic prescription 1/1/18 - 11/30/18 (N)</b>	<b>Number of new antipsychotic prescriptions 1/1/18 - 11/30/18 (N)</b>	<b>Number of new antipsychotic prescriptions that meet the criteria for inclusion in the measure (Denom) (N)</b>
State 6 W	293,048	33,812	186,698	113,639	14,896	6,962	5,310
State 7 NE	213,299	39,839	145,694	113,997	13,291	5,484	3,509
State 8 W	861,540	116,728	604,776	186,322	29,450	15,834	12,619
State 9 W	789,876	85,829	497,095	219,431	33,795	18,954	12,142
State 10 S	870,969	134,210	407,949	298,479	40,215	18,790	13,569
State 11 S	1,264,019	223,218	721,548	515,862	70,721	37,412	29,985
State 12 NE	162,639	38,750	129,709	73,622	9,761	4,243	3,176
State 13 W	1,955,531	233,341	1,261,046	661,313	86,349	41,108	30,657
Mean	1,103,860	187,234	682,486	390,377	53,759	27,872	21,928

\* Cell intentionally left empty

Data Sources: Transformed Medicaid Statistical Information System (T-MSIS) Analytic files (TAF) Research Identifiable Files (RIFs): demographic and eligibility (DE), other services (OT), inpatient (IP) and pharmacy (Rx). Medicare Advantage (MA) encounter files and Medicare Fee-For-Service files: carrier, outpatient (OP), inpatient (IP), MedPAR and prescription drug event (PDE).

An antipsychotic prescription is considered new if the beneficiary did not fill an antipsychotic medication within the previous 120 days (16 weeks).

The inclusion criteria are: 1) beneficiary was continuously enrolled in Medicaid during the 120 days prior to and the 28 days after the new antipsychotic medication fill; 2) beneficiary did not die during the 28 days after the new antipsychotic medication fill; 3) patient did not have an acute inpatient hospitalization during the during the 28 days after the new antipsychotic medication fill.

As shown in Table 2, 37.1 percent of new antipsychotic prescriptions in the analysis were for dual eligible beneficiaries. Over half (51.4 percent) of the new antipsychotic prescriptions were for beneficiaries who qualified for Medicaid due to disability. Nearly half (46.2 percent) of these prescriptions were for beneficiaries between 18 and 45 years of age. White non-Hispanic beneficiaries accounted for about half (52.2 percent) of the new antipsychotic prescriptions, followed by non-Hispanic Black (18.5 percent) and Hispanic beneficiaries (15.6 percent).

Table 2. Demographic Characteristics of the Analytic Sample

<b>*</b>	<b>Number of New Antipsychotic Prescriptions</b>	<b>Percentage</b>
Total (all 13 States)	285,060	100%
Dually Eligible for Medicare	*	*
Medicaid Only	179,323	62.9%
Dual Eligible	105,737	37.1%
Medicaid eligibility category	*	*

*	Number of New Antipsychotic Prescriptions	Percentage
Aged	43,132	15.1%
Disabled	146,598	51.4%
Adult	89,012	31.2%
CHIP/Child/Unknown	6,318	2.2%
Age, Years	*	*
18–34	80,269	28.2%
35–44	51,416	18.0%
45–54	55,680	19.5%
55–64	53,447	18.7%
65–74	22,382	7.9%
75+	21,866	7.7%
Gender	*	*
Female	172,539	60.5%
Male	112,521	39.5%
Race and Ethnicity	*	*
White	148,922	52.2%
Black	52,749	18.5%
Hispanic	44,453	15.6%
Other/Unknown	38,936	13.7%

Data Sources: Transformed Medicaid Statistical Information System (T-MSIS) Analytic files (TAF) Research Identifiable Files (RIFs): demographic and eligibility (DE), other services (OT), inpatient (IP) and pharmacy (Rx). Medicare Advantage (MA) encounter files and Medicare Fee-For-Service files: carrier, outpatient (OP), inpatient (IP), MedPAR and prescription drug event (PDE).

\*Cell intentionally left empty

### 2016 Submission

Of the Medicaid-only and dual-eligible beneficiaries over age 18 years in our sample states, 550,842 beneficiaries had at least one antipsychotic prescription in the measure year. Because the measure looks at whether follow-up occurred after a new antipsychotic prescription (that is, the measure is at the prescription rather than the beneficiary level), the remaining description focuses on the number of prescriptions rather than the number of beneficiaries. There were 3,768,880 antipsychotic prescriptions filled in calendar year 2014 by the Medicaid-only and dual-eligible beneficiaries in our sample; 332,736 of these prescriptions were filled during our measure period (January 1, 2014 through November 30, 2014) and met our definition of a “new antipsychotic prescription.” Of these, 267,831 prescriptions met the continuous enrollment requirements and the inpatient hospitalization and death exclusions.

Table 1. Analytic sample selection (1/1/2014 – 11/30/2014)

*	Total Medicaid-only beneficiaries (N)	Total dual-eligible beneficiaries (N)	Total Medicaid-only and dual-eligible beneficiaries age 18 years and older as of January 1, 2014 (N)	Total antipsychotic prescriptions (N)	Total beneficiaries with any antipsychotic prescriptions (N)	Total new antipsychotic prescriptions <sup>a</sup> (N)	Total new antipsychotic prescriptions, with exclusions <sup>b</sup> (N)
Level of analysis	Beneficiary	Beneficiary	Beneficiary	Prescription	Beneficiary	Prescription	Prescription
<b>Total</b>	<b>19,380,564</b>	<b>3,571,243</b>	<b>12,936,390</b>	<b>3,768,880</b>	<b>550,842</b>	<b>332,736</b>	<b>267,831</b>
State A	771,712	137,575	435,175	79,666	13,163	7,918	5,751
State B	718,183	181,383	583,689	222,802	29,609	17,266	13,150
State C	1,757,408	332,780	844,614	230,905	37,096	22,919	18,705
State D	611,524	93,854	391,053	132,056	18,277	11,247	8,705
State E	2,269,049	318,934	1,439,552	412,981	60,445	36,099	29,545
State F	957,709	246,473	612,290	275,446	38,505	23,063	18,025
State 5	615,772	172,184	371,026	102,246	17,345	10,392	8,640
State G	1,559,258	242,949	1,038,323	274,640	38,217	22,704	17,937
State H	5,896,113	918,391	4,462,679	1,090,381	156,671	92,636	75,644
State I	2,110,460	475,951	1,396,822	599,029	84,835	52,128	41,979
State J	124,816	22,826	55,710	25,877	2,977	1,517	1,141
State K	1,218,466	283,765	746,666	204,438	35,158	22,538	19,269
State 12	168,349	37,646	136,889	33,391	4,948	3,201	2,525
State L	529,276	94,121	390,341	76,330	12,176	8,164	6,083
State M	72,469	12,411	31,561	8,692	1,420	944	732
Mean	1,292,038	238,083	862,426	251,259	36,723	22,182	17,855

Source: Mathematica analysis of 2013–2014 MAX PS, RX, OT, and IP files and Medicare Part B, Part D, Outpatient Department, and MedPAR files.

\* Cell intentionally left empty

<sup>a</sup> An antipsychotic prescription is considered new if the beneficiary was not prescribed any antipsychotic medications within the previous 16 weeks.

<sup>b</sup> Exclusions limit the sample to prescriptions to beneficiaries who (1) were continuously enrolled in Medicaid for 16 weeks before and 4 weeks following the prescription, (2) did not die within 4 weeks following the prescription, and (3) did not have an inpatient stay within 4 weeks of prescription.

Slightly more than half of the prescriptions in the analytic sample were for dual-eligible beneficiaries. Almost a quarter of prescriptions in the analytic sample were filled for beneficiaries ages 18 to 34 (Table 2), whereas only 8.6 percent of prescriptions were filled for beneficiaries ages 65 to 74. Well over half of the prescriptions were for female beneficiaries (58.5 percent). White beneficiaries accounted for more than half of the new antipsychotic prescriptions included in the measure (57.0 percent), followed by black and Hispanic beneficiaries (26.5 and 10.4 percent, respectively).

Table 2. Descriptive statistics of the analytic sample

Characteristic	Number of qualifying prescriptions	Percentage
Total	267,831	100.00
Medicaid beneficiary category	*	*
Medicaid only	132,835	49.60
Dual-eligible	134,996	50.40
Age	*	*

Characteristic	Number of qualifying prescriptions	Percentage
18–34	65,162	24.33
35–44	45,932	17.15
45–54	59,976	22.39
55–64	44,690	16.69
65–74	23,067	8.61
75+	29,004	10.83
Sex	*	*
Female	156,750	58.53
Male	111,081	41.47
Race/ethnicity	*	*
White	152,748	57.03
Black	70,886	26.47
Hispanic	27,838	10.39
Other or unknown race/ethnicity	16,359	6.11
Disability status	*	*
Disabled	180,564	67.42
Not disabled	87,267	32.58

Source: Mathematica analysis of 2013–2014 MAX PS, RX, and IP files and Medicare Part B, Part D, and MedPAR files. Disability status is determined by either Medicaid or Medicare enrollment data indicating disability. For all beneficiaries, we used the monthly Maintenance Assistance Status/Basis of Eligibility (MASBOE) variables in Medicaid enrollment data to determine disability in the month of the new antipsychotic prescription. For dual-eligible beneficiaries, we also used the Original Reason for Entitlement in the Medicare data.

\* Cell intentionally left empty

**[Response Ends]**

**2a.07. 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.**

**[Response Begins]**

**2021 Submission**

Not applicable.

**2016 Submission**

To analyze the impact of the exclusions on measure performance, we computed the denominator, numerator, and state-level performance for variations on the overall sample:

- Exclusions analysis 1: Removing the exclusion rule for beneficiaries who expired during the four-week follow-up period
- Exclusions analysis 2: Removing the exclusion rule for beneficiaries who had an inpatient hospital stay during the four-week follow-up period
- Exclusions analysis 3: Removing both exclusion rules

**[Response Ends]**

**2a.08. List the social risk factors that were available and analyzed.**

*For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.*

**[Response Begins]**

**2021 Submission**

As described in section 1.6, we collected information on the following variables using data extracted from the T-MSIS Analytic files (TAF) Research Identifiable Files (RIFs): dual eligibility for Medicare and Medicaid, Medicaid eligibility

category, age, gender and race/ethnicity. This measure is based on a process that should be carried out for all beneficiaries (except those excluded), so no adjustment for patient mix is necessary. We did collect information about these variables and assessed disparities in performance rate for each group. Those results are described in section 2b5.

#### **2016 Submission**

As described in section 1.6, we collected information on the following variables using data extracted from Medicaid Analytic eXtract (MAX) 2013 and 2014 files: Medicaid eligibility category, age, sex, and race/ethnicity. This measure is based on a process that should be carried out for all beneficiaries (except those excluded), so no adjustment for patient mix is necessary. We did collect information about these variables and assessed disparities in performance rate for each group. Those results are described in section 2b5.

#### **[Response Ends]**

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter “see validity testing section of data elements”; and enter “N/A” for 2a.09 and 2a.10.

#### **2a.09. Select the level of reliability testing conducted.**

*Choose one or both levels.*

#### **[Response Begins]**

Accountable Entity Level (e.g., signal-to-noise analysis)

#### **[Response Ends]**

#### **2a.10. For each level of reliability testing 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.*

#### **[Response Begins]**

#### **2021 Submission**

*Signal-to-noise reliability.* We examined signal-to-noise reliability of NQF 3313 performance using the methodology described by John Adams.<sup>11</sup> This methodology uses a beta-binomial model to assess how well one can confidently distinguish the performance of one reporting entity from another. Conceptually, the beta-binomial model is the ratio of signal to noise. The signal is the proportion of the variability in measured performance that can be explained by real differences in performance across reporting entities (for NQF 3313, the reporting entities are states). The beta-binomial model is an appropriate tool when estimating the reliability of simple pass/fail rate measures, such as NQF 3313. Using this approach, reliability scores can range from 0.0 to 1.0, with a score of 0.0 implying that all variation is attributed to measurement error (noise), and a reliability of 1.0 implying that all variation is caused by a real difference in performance across reporting entities (signal). For NQF 3313, states are the reporting entity at which signal to noise was assessed, as described in the formulas and explanations below.

The formula for signal-to-noise reliability is  $\text{signal-to-noise reliability} = \frac{\sigma^2_{\text{state-to-state}}}{(\sigma^2_{\text{state-to-state}} + \sigma^2_{\text{error}})}$ . Therefore, the team estimated two variances: 1) variance between states ( $\sigma^2_{\text{state-to-state}}$ ), and 2) variance within states ( $\sigma^2_{\text{error}}$ ).

1. *Variance between states:*  $\sigma^2_{\text{state-to-state}} = \frac{\alpha \beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}$

$\alpha$  and  $\beta$  are two shape parameters of the beta-binomial distribution, where  $\alpha > 0$  and  $\beta > 0$ .

1. *Variance within states:*  $\sigma^2_{\text{error}} = \hat{p}(1 - \hat{p})/n$

$\hat{p}$  is the observed rate for the state, and  $n$  is the state-specific denominator for the observed rate (in this context, the number of new antipsychotic prescriptions in the state).

Using Adams’ methodology, the team estimated the reliability for each reporting entity (state), averaging reliability estimates across all reporting entities to produce a point estimate of signal-to-noise reliability (known as the *mean signal-to-noise reliability* value). Mean signal-to-noise reliability shows how well, on average, NQF 3313 differentiates between reporting entity performance for each metric.

Along with the point estimate of mean signal-to-noise reliability, the team also estimated:

1. **The mean, standard deviation, minimum and maximum signal-to-noise reliability for all states.** The standard deviation, minimum and maximum estimates of signal-to-noise reliability provide information about the stability of the reliability results. The narrower the range between the minimum and maximum estimates, the less the signal-to-noise reliability estimate will change due to idiosyncratic features of specific states. Due to the small

number of states included in this analysis, we did not stratify the signal-to-noise reliability estimates by terciles of the denominator.

2. **Key percentiles of the distribution (minimum, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, and maximum) for the state-level signal-to-noise reliability estimates.** Each state’s reliability estimate is a ratio of signal to noise, as described above [that is,  $\sigma^2_{\text{state-to-state}}/(\sigma^2_{\text{state-to-state}}+\sigma^2_{\text{error}})$ ]. Variability between states ( $\sigma^2_{\text{state-to-state}}$ ) is the same for each state, while the specific state error ( $\sigma^2_{\text{error}}$ ) varies. Reliability for each state is an ordinal measure of how well one can determine where a state lies in the distribution across states, with higher estimates indicating better reliability.

**2016 Submission**

The signal-to-noise ratio (SNR) statistic,  $R$  (ranging from 0 to 1), summarizes the proportion of the variation between-state scores that is a result of real differences in the underlying characteristics of a state (such as differences in population demographic characteristics or quality of care provided) as opposed to background-level or random variation (such as measurement or sampling error). If  $R=0$ , there is no variation in the measure across entities, and all observed variation is due to sampling variation. In this case, the measure would not be useful for distinguishing between entities with respect to health care quality. Conversely, if  $R=1$ , all state scores are free of sampling error, and all variation represents real differences between entities in the measure result.

We estimated SNR reliability for the NQF 3313 measure by using a beta-binomial model, which is suitable for binary pass/fail rate measures (Adams 2009). For NQF 3313, the pass/fail designation is defined as the presence or absence of an eligible follow-up visit within a specified time frame after the prescription was filled. The beta-binomial model assumes that the state SNR score is a binomial random variable conditional on the state’s true value, which comes from the beta distribution (ranging from 0 to 1). We calculated SNR reliability in three steps (Adams 2009; Adams 2014; NQF 2016).

First, we calculated state-specific NQF 3313 variance (“noise”) as a function of the measure “passing rate” at the state level, (passed/eligible) and the sample size,  $n$ :

Second, we used version 2.2 of the BETABIN SAS macro written by Wakeling (no date) to fit the beta-binomial model to the NQF 3313 dataset. The macro produced the estimated average pass rate across all providers as well as the Alpha ( $\alpha$ ) and Beta ( $\beta$ ) parameters that describe the shape of the fitted beta-binomial distribution. We calculated the “signal” (between-state variation of the NQF 3313 measure) by using these parameters as follows:

Third, we calculated the SNR reliability as the ratio of the between-state variance and the total variance (i.e., the sum of the between-state and within-state variances) of the NQF 3313 measure rate:

[1] Adams, JL (2009). *The Reliability of Provider Profiling: A Tutorial*. Santa Monica, California: RAND Corporation.

**[Response Ends]**

**2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?**

*For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, [NQF Measure Evaluation Criteria](#)).*

**[Response Begins]**

**2021 Submission**

The signal-to-noise analysis was conducted separately for each of the 13 states in the analysis and overall. NQF 3313 was highly reliable in distinguishing performance between states. The average signal-to-noise reliability was 0.991 and ranged from 0.977 to 0.999 across states (Tables 3a and 3b).

Table 3a. Distribution of State Estimates of Signal-to-Noise Reliability for NQF 3313 - Follow-up Care for Adult Medicare Beneficiaries who are Newly Prescribed an Antipsychotic Medication

Number of States	Mean	Standard Deviation	Minimum	25th percentile	50th percentile	75th percentile	Maximum
13	0.991	0.007	0.977	0.986	0.994	0.998	0.999

Table 3b. State-level estimates of Signal-to-Noise Reliability - NQF 3313 - Follow-up Care for Adult Medicare Beneficiaries who are Newly Prescribed an Antipsychotic Medication



State	Number of New Antipsychotic Prescriptions (Denominator)	Number of New Antipsychotic Prescriptions with an Outpatient Visit within 28 Days (Numerator)	Signal-to-Noise Reliability
Mean (13 states)	21,928	10,047	0.991
State 1 West	33,690	18,052	0.998
State 2 South	6,771	2,428	0.990
State 3 South	104,074	46,260	0.999
State 4 West	5,080	2,426	0.985
State 5 Midwest	24,478	10,991	0.997
State 6 West	5,310	2,273	0.986
State 7 Northeast	3,509	1,647	0.979
State 8 West	12,619	5,505	0.994
State 9 West	12,142	6,384	0.994
State 10 South	13,569	6,877	0.994
State 11 South	29,985	12,064	0.998
State 12 Northeast	3,176	1,865	0.977
State 13 West	30,657	13,849	0.998

Data Sources: Transformed Medicaid Statistical Information System (T-MSIS) Analytic files (TAF) Research Identifiable Files (RIFs): demographic and eligibility (DE), other services (OT), inpatient (IP) and pharmacy (Rx). Medicare Advantage (MA) encounter files and Medicare Fee-For-Service files: carrier, outpatient (OP), inpatient (IP), MedPAR and prescription drug event (PDE).

#### 2016 Submission

**SNR analysis.** The SNR statistic for NQF 3313 was computed separately for each of the 15 states in the sample (Table 3). NQF 3313 was highly reliable in distinguishing performance between states; the average SNR value was 0.98, and it ranged from 0.88 to 0.99 across states. It is important to note that although high reliability is not indicative of high quality health care, it does indicate that the measure may be used to distinguish between states with respect to health care quality.

Table 3. Signal-to-noise reliability for the NQF 3313 measure (N = 15)

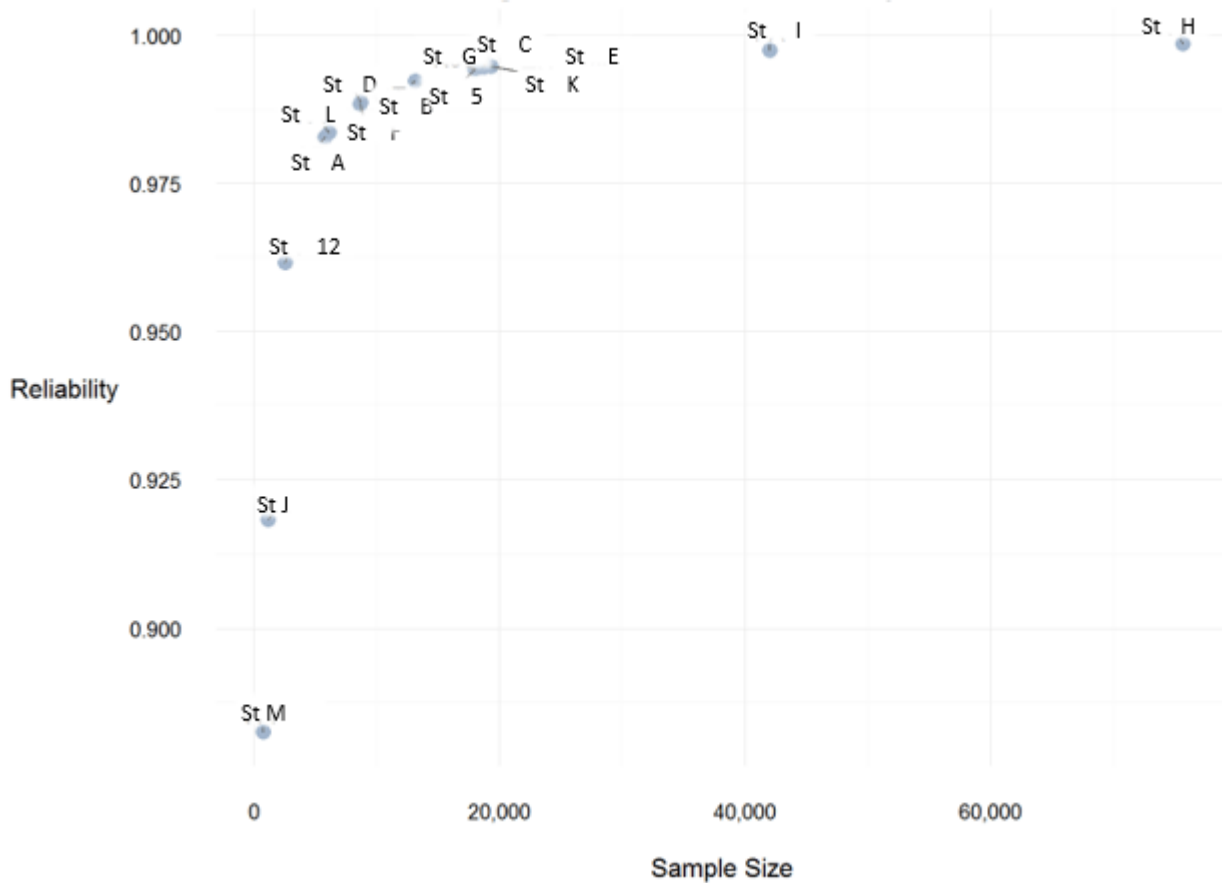
State	Number of new antipsychotic prescriptions (Denominator)	Number of prescriptions with follow-up visit within four weeks (Numerator)	Reliability score
Mean (all states)	17,855	8,719	0.98
State A	5,751	2,558	0.98
State B	13,150	6,234	0.99
State C	18,705	8,473	0.99
State D	8,705	4,986	0.99
State E	29,545	17,177	0.99
State F	18,025	8,883	0.99
State 5	8,640	3,924	0.99
State G	17,937	8,420	0.99
State H	75,644	36,219	0.99
State I	41,979	19,597	0.99
State J	1,141	582	0.92
State K	19,269	8,666	0.99
State 12	2,525	1,401	0.96
State L	6,083	3,225	0.98
State M	732	440	0.88

Source: Mathematica analysis of 2013–2014 MAX PS, RX, OT, and IP files and Medicare Part B, Part D, OPD, and MedPAR files.

Note: The upper boundaries of the SNR statistic for State E, State I, and State H were truncated to 0.99 rather than rounded to 1.00 to reflect the uncertainty in the estimate.

High reliability for NQF 3313 is likely supported by large enough sample sizes at the state level. The average number of newly filled prescriptions per state for NQF 3313 was 17,855 (ranging from 259 to 42,227), and the average number of prescriptions with continuity treatment per state was 8,719 (ranging from 168 to 21,986). In Figure 1, we show the SNR statistics for each state plotted against sample size and demonstrate how reliability increases with sample size. The reliability of all states was above the NQF threshold of 0.70 for acceptable reliability (Measure Testing Task Force Report 2011).

Figure 1. State-level signal-to-noise ratio reliability of NQF 3313 rates



Source: Mathematica analysis of 2013–2014 MAX PS, RX, OT, and IP files and Medicare Part B, Part D, OPD, and MedPAR files.

[Response Ends]

**2a.12. Interpret the results, in terms of how they demonstrate reliability.**

*(In other words, what do the results mean and what are the norms for the test conducted?)*

[Response Begins]

**2021 Submission**

NQF 3313 is rated high for scientific acceptability, based on reliability results. The signal-to-noise analyses showed that the reliability of NQF 3313 is excellent. Although high signal-to-noise reliability is not indicative of high-quality health care, it does indicate that the measure may be used to distinguish between states with respect to health care quality. High reliability for NQF 3313 is likely supported by large enough sample sizes at the state level. The average number of newly filled antipsychotic prescriptions per state for NQF 3313 was 21,928 (ranging from 3,176 to 104,074), and the average number of prescriptions with continuity treatment per state was 10,047 (ranging from 1,647 to 46,260).

[Response Ends]

**2b.01. Select the level of validity testing that was conducted.**

**[Response Begins]**

Accountable Entity Level (e.g. hospitals, clinicians)

Empirical validity testing

**[Response Ends]**

**2b.02. 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.*

**[Response Begins]**

**2021 Submission**

*Convergent validity.* The team examined validity for NQF 3313 through empirical testing. We calculated Spearman correlation coefficients for construct validity using state-level data derived from the Medicaid T-MSIS TAF RIFs as well as HEDIS data, aggregated at the state level. HEDIS data was available for 9 of the 13 states. Given the small sample size (n=9) for this analysis, we calculated Spearman rather than Pearson correlations. Spearman correlations estimate the strength and direction of the monotonic association between two continuous variables; the magnitude of correlation ranges from -1.0 to +1.0. A value of 1.0 indicates a strong positive association (that is, an increase in values of the first variable are associated with an increase in value of the second variable). A value of 0.0 indicates no association between variables. A value of -1.0 indicates a strong negative association (that is, an increase in values of the first variable are associated with a decrease in values of the second variable). The significance of a Spearman 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. P values of less than 0.05 imply that it is unlikely a non-zero coefficient was observed due to chance alone. NQF 3313 represents the percentage of Medicaid beneficiaries aged 18 years and older with new antipsychotic prescriptions who have an outpatient follow-up visit with a provider with prescribing authority within 28 days (four weeks) of the antipsychotic medication prescription fill date. The team used the following research questions to assess NQF 3313:

- Is NQF 3313 correlated with the HEDIS *Follow-Up After Hospitalization for Mental Illness (FUH)*<sup>[1]</sup> measure?
- Is NQF 3313 correlated with the HEDIS *Follow-Up Care for Children Prescribed ADHD Medication (ADD)*<sup>[2]</sup> measure?

The team hypothesized that NQF 3313 will have a positive correlation with both the HEDIS FUH and ADD measures, as the processes assessed in these measures are related or similar to NQF 3313:

- The HEDIS FUH measure assesses a 7-and 30-day follow-up after hospitalization for mental illness. The team hypothesized that high-risk individuals with a mental health disorder who had better access to care following an inpatient event, may also do better with maintaining their care in the outpatient setting.
- The HEDIS ADD measure assesses follow-up care for individuals newly prescribed an attention-deficit/hyperactivity disorder (ADHD) medication, with the evaluation of the initiation phase (1 month) and continuation phase (10 months) after medication is dispensed. Both NQF 3313 and the HEDIS ADD measure assess follow-up for individuals who are newly prescribed medication that warrants close monitoring in outpatient settings.

**2016 Submission**

**Face validity.** To assess face validity, we conducted a survey of TEP members to obtain their assessment of the extent to which the measure's state-level performance scores distinguish good quality from poor quality of care. Of the 11 TEP members who responded to the survey, eight TEP members agreed that the measure distinguishes good quality from poor quality of care. Two TEP members reported a neutral assessment of the performance score's face validity, and one TEP member disagreed with the statement. Overall, this support suggests that the measure has moderate-high face validity for distinguishing the quality of care provided to beneficiaries who are newly prescribed antipsychotic medications.

<sup>[1]</sup> The HEDIS FUH measure assesses the percent of hospital discharge for mental illness or intentional self-harm and had follow-up with a mental health provider. This measure has two indicators: 7-day follow-up and 30-day follow-up.

<sup>[2]</sup> The HEDIS ADD measure assesses follow-up care for children prescribed an ADHD medication. This measure has two indicators: initiation phase and continuation and maintenance phase.

**[Response Ends]**

**2b.03. Provide the statistical results from validity testing.**

*Examples may include correlations or t-test results.*

**[Response Begins]**

**2021 Submission**

Of the 13 states included in this analysis, only 9 had HEDIS FUH and ADD Medicaid data available (State 2 South, State 3 South, State 4 West, State 5 Midwest, State 6 West, State 7 Northeast, State 8 West, State 11 Northeast, State 13 West). As shown in Table 4, NQF 3313 had a moderately positive correlation with the HEDIS FUH measure for the 7-Day and 30-day indicators for follow-up after hospitalization for mental illness ( $r=0.42$  and  $r=0.38$ , respectively). NQF 3313 also showed a moderately positive correlation with the HEDIS ADD measure for the “initiation” and “continuation and maintenance” phase indicators ( $r=0.58$  and  $r=0.53$ , respectively). Due to small sample size of 9 states, it is unsurprising that the results did not reach statistical significance of  $p<0.05$ .

Table 4: Convergent Validity Testing Results: Spearman Correlations Between NQF 3313 and HEDIS FUH 7-Day and 30-Day Follow-up After Hospitalization for Mental Illness and HEDIS ADD Initiation and Continuation Phase following a new prescription of ADHD medication.

<b>Spearman Correlations with NQF 3313</b>	<b>Spearman Correlation</b>	<b>Correlation p Value</b>
FUH: 7-Day Follow-up after hospitalization for mental illness	0.42	0.26
FUH: 30-Day Follow-up after hospitalization for mental illness	0.38	0.31
ADD: Initiation Phase	0.58	0.10
ADD: Continuation & Maintenance Phase	0.53	0.14

**2016 Submission**

For the state-level convergent validity analysis, the PMH team created comparative graphs for visual inspection that compares state-level performance on NQF 3313 to performance on each of the five Core Set measures, respectively. A comparative graph provides suggestive evidence of NQF 3313’s convergent validity if high performance on NQF 3313 relative to other states was correlated with high performance on the Core Set measures relative to other states, and low performance on NQF 3313 relative to other states was correlated with low performance on the Core Set measures relative to other states. The sample of states was too small to conduct these analyses formally (that is, statistically). We find that states with high performance (that is, relatively close to 60.1 percent, the highest state-level performance rate observed during testing) on PMH 1 often had relatively high performance on the following measures:

- Follow-up After Hospitalization for Mental Illness
- Antidepressant Medication Management
- Follow-up Care for Children Prescribed ADHD Medications

The reverse was true as well—states with low performance (that is, relatively close to 45.5 percent, the lowest state-level performance rate observed during testing) on NQF 3313 also had low performance on these three Core Set measures. Table 4 shows states performance on Follow-up After Hospitalization for Mental Illness (FUH-AD) from highest to lowest, and Figure 2 compares states’ performance on NQF 3313 to performance on FUH-AD. State performance on FUH-AD ranged from 45.6 percent in State D to 75.9 percent in State K. Based on the ten states with available data, we found that most states exhibit a roughly positive correlation between the two measures. For example, State 5 has low relative performance on both—45.4 percent on NQF 3313 and 50.3 percent on FUH-AD, but State 12 has relatively high performance on both—55.5 percent on NQF 3313 and 75.4 percent on FUH-AD. State K and State D are the exceptions, the latter of which is an outlier in almost all cases.

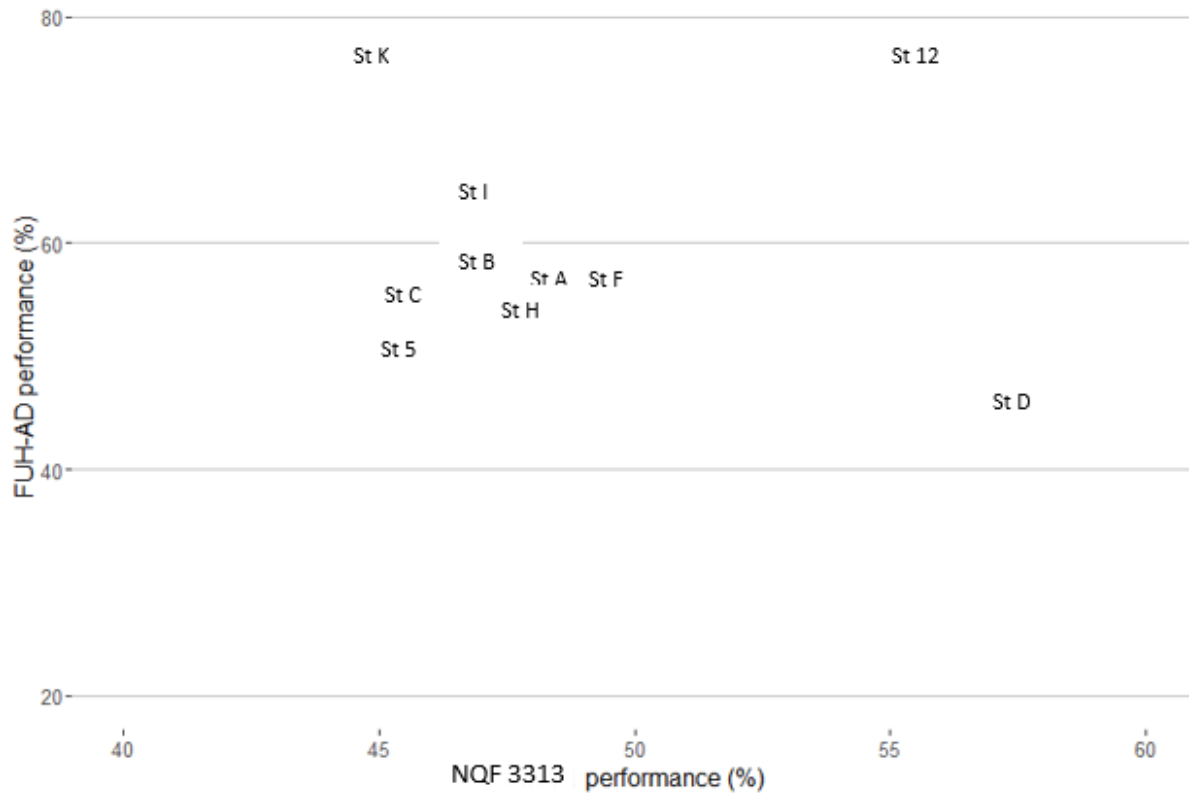
Table 4. State performance on Follow-up After Hospitalization for Mental Illness (FUH-AD)

<b>State</b>	<b>FUH-AD performance (%)</b>
State K	75.9
State 12	75.4
State I	64.2
State B	57.5
State F	57.1
State A	56.7
State C	55.9
State H	55.3
State 5	50.3
State D	45.6

Source: State-reported performance rates for Adult Core Set measures, FFY2015.

Note: State E, State G, State J, State L, and State M did not report state performance rates for the FUH-AD measure.

**Figure 2. State Performance on NQF 3313 and Follow-up After Hospitalization for Mental Illness (FUH-AD)**



Source: Mathematica analysis of state-level performance on Adult Core Set measures, FFY2015 and state-level performance of NQF 3313 using data from 2013–2014 MAX PS, RX, OT, and IP files, and Medicare Part B, Part D, OPD, and MedPAR files.

Table 5 shows state performance on Antidepressant Medication Management (AMM-AD) from highest to lowest, and Figure 3 compares states’ performance on NQF 3313 to performance on AMM-AD. State performance on AMM-AD ranges from 33.3 percent in State A to 69.3 percent in State 12, and most of the states appear clustered around 50 percent performance on AMM-AD and about 48 percent performance on NQF 3313. While State D again appears to be somewhat of an outlier, states with higher relative performance on NQF 3313 often have higher relative performance on AMM-AD.

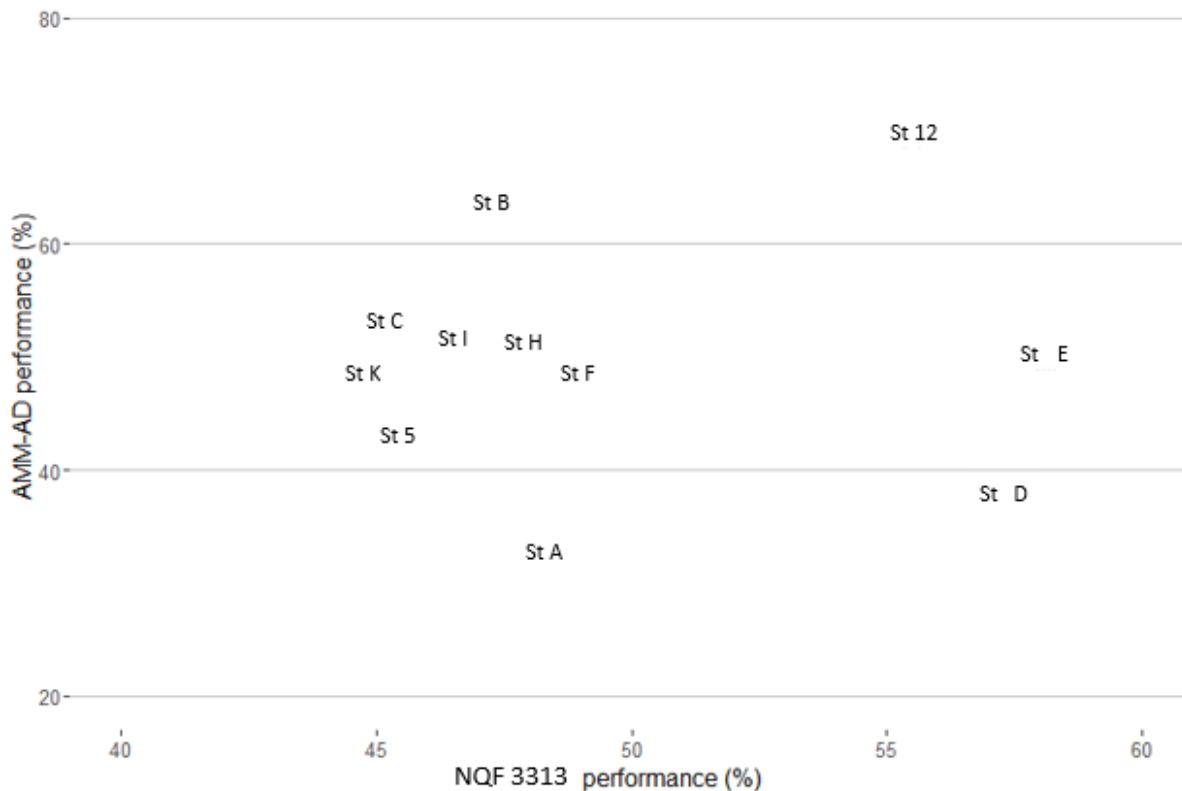
**Table 5. State performance on Antidepressant Medication Management (AMM-AD)**

State	AMM-AD performance (%)
State 12	69.3
State B	63.9
State C	53.5
State I	51.9
State H	51.0
State E	49.7
State F	48.8
State K	48.6
State 5	43.0
State D	37.9
State A	33.3

Source: State-reported performance rates for Adult Core Set measures, FFY2015.

Note: State G, State J, State L, and State M did not report state performance rates for the AMM-AD measure.

**Figure 3. State Performance on NQF 3313 and Antidepressant Medication Management (AMM-AD)**



Source: Mathematica analysis of state-level performance on Adult Core Set measures, FFY2015 and state-level performance of NQF 3313 using data from 2013–2014 MAX PS, RX, OT, and IP files, and Medicare Part B, Part D, OPD, and MedPAR files.

Table 6 shows state performance on Follow-up Care for Children Prescribed ADHD Medication (ADD-CH) from highest to lowest, and Figure 4 compares states’ performance on NQF 3313 to (ADD-CH). ADD-CH ranges from 27.9 percent in State I to 66.7 percent in State 12. Although State E and State D are again moderate outliers, there appears to be a somewhat positive correlation between NQF 3313 and ADD-AD. State 12 has high relative performance on both measures—55.5 percent on NQF 3313 and 66.7 percent on ADD-CH. By contrast, State I has low relative performance on both measures – 46.7 percent on NQF 3313 and 27.9 percent on ADD-CH.

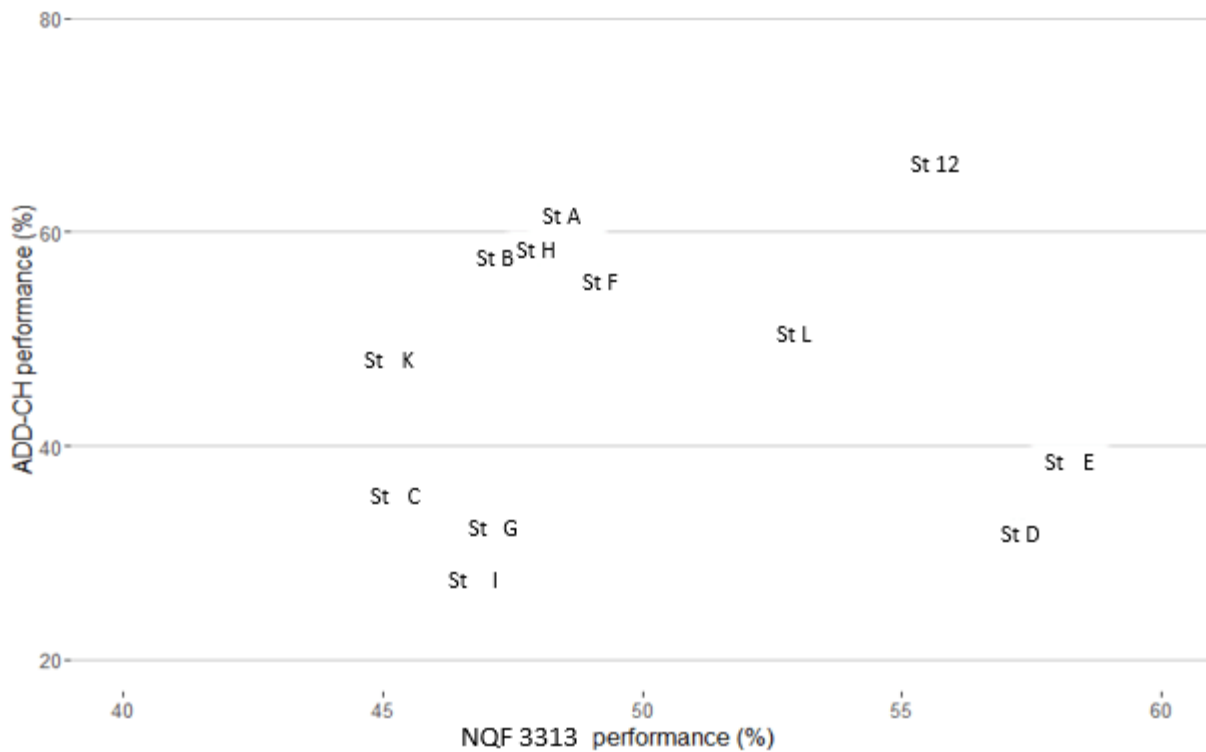
Table 6. State performance on Follow-up Care for Children Prescribed ADHD Medication (ADD-CH)

State	ADD-CH performance (%)
State 12	66.7
State A	61.6
State H	57.8
State B	57.2
State F	56
State L	50.6
State K	47.8
State E	38.9
State C	35.5
State D	32.7
State G	32.5
State I	27.9

Source: State-reported performance rates for Child Core Set measures, FFY2015.

Note: State 5, State J, and State M did not report state performance rates for the ADD-CH measure

**Figure 4. State Performance on NQF 3313 and Follow-up Care for Children Prescribed ADHD Medication (ADD-CH)**



Source: Mathematica analysis of state-level performance on Child Core Set measures, FFY2015 and state-level performance of NQF 3313 using data from 2013–2014 MAX PS, RX, OT, and IP files, and Medicare Part B, Part D, OPD, and MedPAR files.

We did not find a relationship in state performance between NQF 3313 and Adherence to Antipsychotics for Individuals with Schizophrenia (SAA-AD) or Annual Monitoring for Patients on Persistent Medications (MPM-AD). There was very little variation in state performance on these measures, which could explain these findings. State level SAA-AD measure performance ranged from 59.2 percent in State A to 71.7 percent in State I (to provide context, performance on the AMM-AD measure ranged from 33.3 percent in State A to 69.3 percent in State 12). There was even less variation by state on the Medication Monitoring measure (MPM-AD); all states had very high performance (over 87 percent). Where there is little state variation in performance on a Core Set measure, there is minimal leverage with which to compare it to NQF 3313, and therefore difficult to determine whether there is a correlation between the two.

**[Response Ends]**

**2b.04. Provide 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?)**

**[Response Begins]**

**2021 Submission**

NQF 3313 is rated moderate for validity. Data gathered from the project's 2016 TEP showed that nearly 75% of respondents agreed the measure's performance scores are able to distinguish between good quality and poor quality of care for Medicaid beneficiaries newly prescribed antipsychotic medications; and 18% were neutral.

Additionally, state-level performance for NQF 3313 demonstrated a positive association with several measures of similar concepts:

- NQF 3313 had a moderate, positive correlation with the HEDIS FUH measure (7-day follow-up:  $r=0.42$ , 30-day follow-up:  $r=0.38$ ).
- NQF 3313 also showed a moderate, positive correlation with the HEDIS ADD measure (initiation phase:  $r=0.58$ , continuation and maintenance phase:  $r=0.53$ ).

Due to the small sample size of  $N=9$  for these analyses, it is unsurprising that the additional exploratory convergent validity analysis of NQF 3313 with both the HEDIS FUH and HEDIS ADD measures were not statistically significant with a  $p$ -value  $<0.05$ . The differences in measure intent, as well as target population, may also explain the moderate strength in

correlation. For example, the FUH measure is focused on ensuring care coordination for the mental health population from the inpatient setting to the outpatient setting, rather than outpatient only, and the ADD measure is focused on medication management for the child, rather than adult, population.

**2016 Submission**

NQF 3313 is rated moderate for validity. Most respondents from the project's TEP assessed the measure's performance scores as able to distinguish between good quality and poor quality of care for Medicaid beneficiaries newly prescribed antipsychotic medications. Additionally, state-level NQF 3313 rates demonstrated association with several related state-level rates of measures of similar concepts.

**[Response Ends]**

**2b.05. 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 in Importance to Measure and Report: Gap in Care/Disparities.*

**[Response Begins]**

**2021 Submission**

To demonstrate meaningful differences in performance, the team calculated the inter-quartile range (IQR). The IQR provides a measure of the dispersion of performance. The IQR can be interpreted as the difference between the 25<sup>th</sup> and 75<sup>th</sup> percentile on a measure.

To determine if this difference is statistically significant, the team used an independent sample t-test of the performance difference between two randomly selected states: one with performance at or below the 25<sup>th</sup> percentile the other with performance at or above 75<sup>th</sup> percentile. This method takes into account the sample size, difference in performance rates and the variance of the performance rates. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than 0.05, then the two states' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected states, one state below the 25<sup>th</sup> percentile and another state above the 75<sup>th</sup> percentile of performance. We used these two states as examples of measures entities. However, the method can be used for comparison of any two measured entities.

**[Response Ends]**

**2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.**

*Examples may include 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.*

**[Response Begins]**

**2021 Submission**

We found that performance rates across the thirteen states covered a wide range with meaningful variation. Specifically, state-level performance rates ranged from 35.9 percent to 58.7 percent, with an average of 45.8 percent and a standard deviation of 5.0 percent. Seven of the thirteen states (53.8 percent) had performance rates and 95 percent confidence intervals that were below the overall average.

The results of the t-test comparing a randomly selected state above the 75<sup>th</sup> percentile and a randomly selected state below the 25<sup>th</sup> percentile was statistically significant at  $p < 0.001$ . This indicates that the difference in performance rates between the two states likely due to a true difference in performance and unlikely due to chance.

**2016 Submission**

We found that measure rates across the fifteen states covered a wide range with meaningful variation. Specifically, the measure rate ranged from 44.5 percent to 60.1 percent with a mean of 50.2 percent and standard deviation of 5.29 percent. When looking into state-specific measure rate, eight of the 15 states (53.3 percent) exhibited significantly lower measure rates than the average performance, with their 95 percent confidence intervals under the overall performance rate.

**[Response Ends]**



**2b.07. Provide 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.**

*In other words, what do the results mean in terms of statistical and meaningful differences?*

**[Response Begins]**

**2021 Submission**

These findings suggest room for improvement in follow-up after prescription of new antipsychotic medications in the states included in testing. Nine of the 13 states (69.2 percent) had performance rates below 50 percent, indicating that less than half of Medicaid beneficiaries had a follow-up visit within 28 days after a new antipsychotic prescription. None of the 13 states had a performance rate above 59 percent for a follow-up visit after a new antipsychotic prescription. Additionally, the interquartile range (difference between the 75<sup>th</sup> and 25<sup>th</sup> percentiles) was 7.1 percent, and this difference is statistically significant at  $p < 0.001$ . None of the states had performance rates above 59 percent.

**2016 Submission**

Five states showed significantly higher measure rates than the average performance, and two states had performance statistically not distinguishable from the average performance. Overall, these findings indicate both statistically significant and practically meaningful differences in NQF 3313 performance. Interpretation of these comparisons should be tempered by the fact that only 15 states were included in this analysis; the mean rate for the entire nation could be different. Moreover, when discussing room for improvement, states that are not statistically different from or even above the mean rate of 50.2 percent have room for improvement. Only one state reached a rate of 60 percent for follow-up care after a new antipsychotic prescription.

**[Response Ends]**

**2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.**

*Describe the steps—do not just name a method; what statistical analysis was used.*

**[Response Begins]**

**2021 Submission**

*State selection for the analysis.* The team used multiple methods, including Data Quality Atlas (DQA) to identify states with sufficient data quality for inclusion in these analyses. Of the fifteen states included in the prior analysis, four had sufficient data quality to be included in this submission. We then examined the T-MSIS TAF RIFs 2017 data (September 1, 2017 through December 31, 2018) and calendar year 2018 identify states with minimal missing data for the key variables to identify the numerator and denominator for NQF 3313; 1) National Drug Codes (NDCs) and fill date in the prescription drug (RX) file; 2) procedure and diagnosis codes in the other services (this file includes outpatient claims) and inpatient (IP) files. We identified thirteen states that varied in size and geography for inclusion in these analyses.

**2016 Submission**

The extent of missing data was assessed using the MAX validation and anomaly tables (cited below).

**[Response Ends]**

**2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.**

*For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each.*

**[Response Begins]**

**2021 Submission**

Selection methods described above resulted in negligible missing T-MSIS data for the required data elements to calculate the measure for the states in the study sample such as dates of services, date of birth, Medicaid eligibility, prescription fill date, National Drug Code (NDC), and type of service.

**2016 Submission**

The vast majority of the data elements required to calculate the measure—dates of service, date of birth, Medicaid eligibility, prescription fill date, National Drug Code (NDC), and type of service—have negligible missingness in MAX data for the states in the study sample. The one data element that is missing information for more than 5 percent of claims is the field for CPT/HCPCS code.

Table 9 below contains eligibility-related missingness information from 2013 and where available, 2014 MAX data.

Overall, well over 95 percent of MAX claims can be matched to eligibility information.

Table 9. Percent of claims missing corresponding Medicaid eligibility information

State	Year	% with Claims and Missing Medicaid Eligibility (Excludes S-CHIP Only)	IP: % Missing Eligibility and > \$0 Paid (Excludes S-CHIP Only)	LT: % Missing Eligibility and > \$0 Paid (Excludes S-CHIP Only)	OT: % Missing Eligibility and > \$0 Paid (Excludes S-CHIP Only)
State A	2013	2.46	5.33	0.29	0.44
State B	2013	0.27	0.22	0.07	0.18
State C	2013	0.96	0.12	0.02	0.17
*	2014	0.85	0.07	0.01	0.16
State D	2013	0.17	0.14	0.04	0.01
*	2014	0.08	0.04	0.02	0.00
State E	2013	4.08	1.59	0.41	0.38
*	2014	1.50	0.94	0.46	0.10
State 5	2013	2.03	0.14	0.01	0.97
*	2014	0.35	0.21	0.02	0.07
State 4	2013	0.11	0.54	0.02	0.04
*	2014	0.23	0.28	0.02	0.10
State G	2013	0.55	0.20	0.42	0.21
*	2014	0.55	0.24	0.33	0.21
State H	2013	0.07	0.23	0.21	0.00
State I	2013	2.82	1.01	0.47	0.08
*	2014	3.77	0.94	0.16	0.31
State J	2013	0.01	0.00	0.00	0.01
*	2014	0.01	0.00	0.00	0.00
State K	2013	0.39	0.00	0.00	0.03
*	2014	0.56	0.00	0.00	0.03
State 12	2013	0.53	0.93	0.44	0.17
*	2014	0.22	0.41	0.29	0.04
State L	2013	3.60	0.14	0.01	0.19
*	2014	0.09	0.05	0.02	0.01
State M	2013	0.57	1.36	0.23	0.20
*	2014	0.75	1.65	0.32	0.18

\* Cell intentionally left empty

Source: MAX validation tables. Available at the following URL: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAX-Validation-Reports.html?DLsort=0&DLEntries=10&DLPage=1&DLsortDir=ascending>

Table 10 below contains missingness related to claims data elements used in the calculation of the measure.

Table 10. Percent of FFS claims missing data elements used in the calculation of the measure

State	Year	% IP Stays (MAX TOS=01)	OT % with HCPCS or CPT-4 Code
State A	2013	100.0	100.0
State B	2013	98.6	82.3
State C	2013	100.0	100.0
*	2014	100.0	100.0
State D	2013	100.0	94.9
*	2014	100.0	94.3
State E	2013	100.0	95.4
*	2014	100.0	93.1
State F	2013	100.0	100.0
*	2014	100.0	100.0
State 5	2013	99.3	100.0
*	2014	99.3	100.0
State G	2013	99.2	100.0
State H	2013	100.0	26.3
State I	2013	99.6	100.0
*	2014	99.6	100.0
State J	2013	100.0	99.9
*	2014	100	99.9
State K	2013	0.0	100.0
*	2014	0.0	100.0
State 12	2013	99.8	100.0
*	2014	99.9	100.0
State L	2013	100.0	100.0
*	2014	100.0	100.0
State M	2013	97.7	100.0
*	2014	97.7	100.0

\* Cell intentionally left empty

Source: MAX anomaly tables. Available at the following URL: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html>.

Note: State K had virtually no inpatient FFS claims in 2013 or 2014.

Type of service is necessary for identifying whether the beneficiary had an inpatient, prescription, other service, or SNF stay, and as a proxy for prescribing authority. This data element is virtually never missing in MAX data because it is used for binning claims into different files (inpatient, prescription, other services, etc.); see, for example in Table 10 – almost every FFS inpatient (IP) claim has a type of service of “01” (“inpatient”). The one exception is State K, which has virtually no inpatient FFS claims in 2013 or 2014.

The measure hinges on identifying beneficiaries taking a new antipsychotic prescription during the measurement year, which is identified using NDC and prescription fill date. In no states are claims missing NDC. Further, prescription fill date is a required field to be included in MAX data and therefore would not be missing for any claims. Similarly, dates of service are required fields for MAX IP and OT claims and therefore are not missing.

Finally, Table 11 shows missingness of date of birth, sex, or race, which we used to compute measure performance by subgroups. A very low percentage of Medicaid enrollees have missing date of birth or sex. There are relatively high levels of enrollees missing race, so tabulations of measure performance by race will only be possible for a subset of the population. Date of death is only populated for beneficiaries who expired, so it will be missing—by design—in most cases.

Table 11. Percent of Medicaid enrollees with missing date of birth, sex, or race

State	MAX year	Percent of enrollees missing date of birth	Percent of enrollees with missing sex	Percent of enrollees with missing race
State A	2012	0.0	0.0	14.6
State B	2012	0.0	0.0	0.0
State C	2013	0.0	0.0	11.1
State D	2013	0.0	0.0	43.8
State E	2012	0.0	0.0	10.7
State F	2013	0.0	0.0	6.1
State 5	2012	0.0	0.0	4.1
State G	2012	0.0	0.0	28.0
State H	2013	1.3	1.0	7.7
State I	2013	0.0	0.0	12.3
State J	2013	0.0	0.0	0.0
State K	2013	0.0	0.0	10.9
State 12	2013	0.0	0.0	26.2
State L	2013	0.0	0.0	1.5
State M	2013	0.0	0.0	14.4

Source: MAX anomaly tables. Available at the following URL: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html>.

**[Response Ends]**

**2b.10. Provide 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 non-responders), and how the specified handling of missing data minimizes bias.**

*In other words, 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 was conducted, justify the selected approach for missing data.*

**[Response Begins]**

**2021 Submission**

Due to the state selection process used for this analysis, the likelihood of missing data is negligible and would not contribute to systematic bias.

**2016 Submission**

Given the relatively small amount of missing information, we don't believe there is any systematic bias. In addition, states implementing the measure will likely have even less missing data because they will be able to account for their state-specific codes when constructing the measure.

**[Response Ends]**

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

**2b.11. Indicate whether there is more than one set of specifications for this measure.**

**[Response Begins]**

No, there is only one set of specifications for this measure

**[Response Ends]**

**2b.12. 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. Indicate what statistical analysis was used.*

**[Response Begins]**

**[Response Ends]**

**2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.**

*Examples may include correlation, and/or rank order.*

**[Response Begins]**

**[Response Ends]**

**2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.**

*In other words, what do the results mean and what are the norms for the test conducted.*

**[Response Begins]**

**[Response Ends]**

**2b.15. Indicate whether the measure uses exclusions.**

**[Response Begins]**

Yes, the measure uses exclusions.

**[Response Ends]**

**2b.16. Describe the method of testing exclusions and what was tested.**

*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?*

**[Response Begins]**

**2021 Submission**

Refer to prior testing.

**2016 Submission**

We tested the impact of the two exclusion rules to the NQF 3313 measure rates proposed by the Clinical Advisory Workgroup:

1. **Hospitalization** - Medicaid beneficiaries with an acute inpatient admission during the four-week follow-up period after prescription of an antipsychotic medication; and
2. **Death** - Patients who expired within four weeks of new prescription date.

The current NQF 3313 definition excludes both cases in the denominator definition. Based on this, we tested the measure rate in the following four scenarios:

1. NQF 3313 measure specification (pmh1\_current): the current measure specification (i.e., exclude both cases from denominator)

2. Exclusions analysis 1 (pmh1\_nodeath): do not apply the death exclusion rule (i.e., only exclude those beneficiaries with an acute inpatient admission during the four-week follow-up period after prescription of an antipsychotic medication)
3. Exclusions analysis 2 (pmh1\_nohosp): do not apply the hospitalization exclusion rule (i.e., only exclude those who expired within four weeks of new prescription date)

Exclusions analysis 3 (pmh1\_noexcl): neither exclusion rule applied

**[Response Ends]**

**2b.17. Provide 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.*

**[Response Begins]**

**2021 Submission**

Refer to prior testing.

**2016 Submission**

**Denominator sizes.** The exclusion rule has small influence on the eligible denominator size. The measure with both exclusion rules applied (i.e., current measure specs) includes 94.7 percent (or 267831 out of 282905) of the denominator as opposed to the measure not applying any exclusion rule. In general, the hospitalization rule excluded more prescriptions (95.2 percent) from the denominator than the death rule (99.4 percent). This was the case across all states other than State H, for which the death rule excluded more prescriptions from the denominator than the hospital rule.

**Measure performance.** Removing either or both of the hospitalization and death exclusion rules result in a slight change on the overall NQF 3313 measure rates (Table 7). Moreover, variations on measure results are small across all states, and performance rankings are (almost) the same between all pairs, as evidenced by Spearman rank correlations extremely close or equal to 1 (Table 8).

Table 7. NQF 3313 performance rates (all states), with and without exclusions applied

Exclusions applied	NQF 3313 rate (all states)
Exclude if inpatient hospitalization or expired during follow-up period	48.8%
Exclude if inpatient hospitalization during follow-up period	48.7%
Exclude if expired during follow-up period	49.1%
No exclusions	49.0%

Source: Mathematica analysis of 2013–2014 MAX PS, RX, and IP files and Medicare Part B, Part D, and MedPAR files.

Table 8. Spearman rank correlations for NQF 3313 rates with and without exclusions applied

*	pmh1_nodeath	pmh1_nohosp	pmh1_noexcl	pmh1_current
pmh1_nodeath	1	.997	.997	1
pmh1_nohosp	.997	1	1	.997
pmh1_noexcl	.997	1	1	.997
pmh1_current	1	.997	.997	1

Source: Mathematica analysis of 2013–2014 MAX PS, RX, and IP files and Medicare Part B, Part D, and MedPAR files.

\* Cell intentionally left empty

**[Response Ends]**

**2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.**

*In other words, 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.*

**[Response Begins]**

**2021 Submission**

Refer to prior testing.

**2016 Submission**

Although the exclusions have limited impact on the denominator size and measure performance, we maintained the exclusions in the measure specifications, given that (1) the clinical advisory workgroup contributed to the development of these exclusions and the Technical Expert Panel supported the exclusions as appropriate during their review of testing results, and (2) the exclusions are relatively straightforward to calculate and are not expected to impose significant additional burden in measure implementation.

**[Response Ends]**

**2b.19. Check all methods used to address risk factors.**

**[Response Begins]**

No risk adjustment or stratification

**[Response Ends]**

**2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.**

**[Response Begins]**

**[Response Ends]**

**2b.21. 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 (i.e., case mix) is not needed to achieve fair comparisons across measured entities.**

**[Response Begins]**

**2021 Submission**

Not Applicable—No risk adjustment or stratification.

**2016 Submission**

Not Applicable - No risk adjustment or stratification.

**[Response Ends]**

**2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.**

**[Response Begins]**

**[Response Ends]**

**2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.**

*Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of  $p < 0.10$  or other statistical tests; correlation of  $x$  or higher. Patient factors should be present at the start of care, if applicable. Also discuss any “ordering” of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).*

**[Response Begins]**

**[Response Ends]**

**2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.**

**[Response Begins]**

**[Response Ends]**

**2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.**

*Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.*

[Response Begins]

[Response Ends]

**2b.26. 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 control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter “N/A” for questions about the statistical risk model discrimination and calibration statistics.**

*Validation testing should be conducted in a data set that is separate from the one used to develop the model.*

[Response Begins]

[Response Ends]

**2b.27. Provide risk model discrimination statistics.**

*For example, provide c-statistics or R-squared values.*

[Response Begins]

[Response Ends]

**2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).**

[Response Begins]

**2021 Submission**

Not Applicable—No risk adjustment or stratification.

**2016 Submission**

Not Applicable - No risk adjustment or stratification.

[Response Ends]

**2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.**

*The preferred file format is .png, but most image formats are acceptable.*

[Response Begins]

[Response Ends]

**2b.30. Provide the results of the risk stratification analysis.**

[Response Begins]

[Response Ends]

**2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).**



*In other words, what do the results mean and what are the norms for the test conducted?*

**[Response Begins]**

**[Response Ends]**

**2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.**

*Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.*

**[Response Begins]**

**[Response Ends]**

## Criteria 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.

---

### 3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

**[Response Begins]**

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

**[Response Ends]**

### 3.02. Detail to what extent the specified data elements are available electronically in defined fields.

*In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.*

**[Response Begins]**

ALL data elements are in defined fields in electronic claims

**[Response Ends]**

### 3.03. 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 data elements not from electronic sources.

**[Response Begins]**

Not applicable.

**[Response Ends]**

### 3.04. Describe any efforts to develop an eCQM.

**[Response Begins]**

Not applicable.

**[Response Ends]**

### 3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

**[Response Begins]**

We identified 13 states ranging in size and geographical location across the U.S. We used a variety of tools, including Data Quality Atlas (DQA) to identify states with sufficient data quality for these analyses, focusing on the germane data files (e.g., pharmacy claims) and key fields within those data files (national drug code [NDC] and fill date on pharmacy claims).

**[Response Ends]**

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

### 3.07. Detail 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),

**Attach the fee schedule here, if applicable.**

**[Response Begins]**

Not applicable, no fees or licensing are currently required.

**[Response Ends]**

## Criteria 4: Use and Usability

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.

---

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

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 demonstrating performance improvement.

### 4a.01. Check all current uses. For each current use checked, please provide:

**Name of program and sponsor**

**URL**

**Purpose**

**Geographic area and number and percentage of accountable entities and patients included**

**Level of measurement and setting**

#### [Response Begins]

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

#### **[Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Please Explain]**

MEDICAID INNOVATION ACCELERATOR PROGRAM: In July 2014, the Centers for Medicare & Medicaid Services (CMS) launched the Medicaid Innovation Accelerator Program (IAP), a collaborative between the Center for Medicaid and CHIP Services (CMCS) and the Center for Medicare & Medicaid Innovation (CMMI). The goal of IAP was to improve the health and health care of Medicaid beneficiaries and to reduce costs by supporting states' ongoing payment and delivery system reforms. Medicaid IAP supported state Medicaid agencies to build capacity in key areas by offering targeted technical support, tool development, and cross-state learning opportunities. The IAP covered all 50 states. The measure was developed and published for optional reporting by states.

The IAP's Physical and Mental Health Integration area worked with states to support integrated care approaches that have been shown to improve health outcomes for individuals with behavioral health conditions. Effective integrated care can also enhance patient engagement and activation, which has been shown to be associated with increased treatment adherence, improved patient satisfaction, better quality of life, and increased mental and physical health.

In 2018, CMS announced the Section 1115 Demonstration for Serious Mental Illness (SMI) and Serious Emotional Disturbance (SED) created under the authority of section 1115(a) of the Social Security Act to enable states to demonstrate and test flexibilities to improve the continuum of care for beneficiaries with SMI/SED. There are 28 states that are approved for this demonstration. States may use NQF 3313 to assess the continuum of mental health care for antipsychotic use. Seven states will likely report NQF 3313 in the near future: District of Columbia, Idaho, Indiana, Oklahoma, Utah, Vermont and Washington.

Quality Improvement (Internal to the specific organization)

#### **[Quality Improvement (Internal to the specific organization) Please Explain]**

MEDICAID INNOVATION ACCELERATOR PROGRAM: In July 2014, the Centers for Medicare & Medicaid Services (CMS) launched the Medicaid Innovation Accelerator Program (IAP), a collaborative between the Center for Medicaid and CHIP Services (CMCS) and the Center for Medicare & Medicaid Innovation (CMMI). The goal of IAP was to improve the health and health care of Medicaid beneficiaries and to reduce costs by supporting states' ongoing payment and delivery system reforms. Medicaid IAP supported state Medicaid agencies to build capacity in key areas by offering targeted technical support, tool development, and cross-state learning opportunities. The IAP covered all 50 states. The measure was developed and published for optional reporting by states.

The IAP's Physical and Mental Health Integration area worked with states to support integrated care approaches that have been shown to improve health outcomes for individuals with behavioral health conditions. Effective integrated care

can also enhance patient engagement and activation, which has been shown to be associated with increased treatment adherence, improved patient satisfaction, better quality of life, and increased mental and physical health. In 2018, CMS announced the Section 1115 Demonstration for Serious Mental Illness (SMI) and Serious Emotional Disturbance (SED) created under the authority of section 1115(a) of the Social Security Act to enable states to demonstrate and test flexibilities to improve the continuum of care for beneficiaries with SMI/SED. There are 28 states that are approved for this demonstration. States may use NQF 3313 to assess the continuum of mental health care for antipsychotic use. Seven states will likely report NQF 3313 in the near future: District of Columbia, Idaho, Indiana, Oklahoma, Utah, Vermont and Washington.

**[Response Ends]**

**4a.02. Check all planned uses.**

**[Response Begins]**

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

Quality Improvement (internal to the specific organization)

**[Response Ends]**

**4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.**

*For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?*

**[Response Begins]**

CMS is considering implementation options for this measure. There are no identified barriers to implementation in a publicly reporting or accountability application.

**[Response Ends]**

**4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.**

*A 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.*

**[Response Begins]**

CMS is developing measures to improve the quality of care of the following Medicaid populations served by CMS's Innovation Accelerator Program:

- People eligible for both Medicare and Medicaid, or "Dual-eligible beneficiaries"
  - People receiving long-term services and supports (LTSS) through managed care organizations
  - People with substance use disorders; beneficiaries with complex care needs and high costs; beneficiaries with physical and mental health needs; or Medicaid beneficiaries who receive LTSS in the community
- This measure is intended for voluntary use by states to monitor and improve the quality of care provided for Medicaid beneficiaries with physical and mental health integration needs. States may choose to begin implementing the measures based on their programmatic needs.

**[Response Ends]**

**4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**

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

**[Response Begins]**

Assistance with measure implementation was provided to interested stakeholders during a CMS-sponsored webinar on July 10, 2019.

**[Response Ends]**

**4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

**[Response Begins]**

During a webinar on July 10, 2019, participants were given information on measure implementation, including steps to calculate measure performance. Participants were given the opportunity to have their questions answered by the steward and developer.

**[Response Ends]**

**4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.**

**[Response Begins]**

No feedback has been received on measure performance and implementation from measure implementers.

**[Response Ends]**

**4a.08. Summarize the feedback obtained from those being measured.**

**[Response Begins]**

We have not received feedback on measure performance and implementation from measure implementers.

**[Response Ends]**

**4a.09. Summarize the feedback obtained from other users.**

**[Response Begins]**

We have not received feedback from other users.

**[Response Ends]**

**4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**

**[Response Begins]**

Not applicable.

**[Response Ends]**

**4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. 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.**

**[Response Begins]**

Adoption of this performance measure has the potential to improve the quality of care for Medicaid beneficiaries newly prescribed with antipsychotic medications. Follow-up care after a newly prescribed antipsychotic medication can be important for monitoring side effects, providing patient education, and adjusting dosage or medications as needed. The results of the measure testing efforts for this submission show room for improvement, as state-level performance rates ranged from 35.9 percent to 58.7 percent. For nine of the 13 states included in this analysis, less than half of Medicaid

beneficiaries who received a new antipsychotic medication had an outpatient follow-up visit within 28 days. Additionally, state-level performance in 2014 and 2018 was similar for the two states used in both measure testing submissions (3.2 percentage point improvement for one state and -0.5 percentage point decline for the other state). The measure intent is to ensure timely follow-up visits to monitor new antipsychotic medications for beneficiaries. There are several strategies that may be used to improve follow-up rates such as care management or coordination strategies to encourage beneficiaries to engage in follow-up care and improved communication between providers and beneficiaries.

**[Response Ends]**

**4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.**

**[Response Begins]**

To date, there have been no unexpected findings identified during use of this measure.

**[Response Ends]**

**4b.03. Explain any unexpected benefits realized from implementation of this measure.**

**[Response Begins]**

To date, there have been no unexpected findings identified during use of this measure.

**[Response Ends]**

## Criteria 5: Related and 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.

---

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

### **5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).**

*(Can search and select measures.)*

**[Response Begins]**

0108: Follow-Up Care for Children Prescribed ADHD Medication (ADD)

3539e: Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting

**[Response Ends]**

### **5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).**

*(Can search and select measures.)*

**[Response Begins]**

**[Response Ends]**

### **5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.**

**[Response Begins]**

Not applicable. There are no other non-NQF-endorsed measures that conceptually address the same measure focus and same target population.

**[Response Ends]**

### **5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.**

**[Response Begins]**

Yes

**[Response Ends]**

### **5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

**[Response Begins]**

This measure differs from NQF 0108 in that it focuses on adults rather children, and on antipsychotic medications rather than ADHD medications. The measures are completely harmonized to the extent possible, with the same follow-up period and look-back period to establish a “new prescription.”

While NQF 3539e and this measure have similar focus on antipsychotic prescribing, this measure differs from NQF 3539e in their target populations and intent. NQF 3313’s focus population is Medicaid beneficiaries aged 18 years and older prescribed a new antipsychotic medication in an outpatient setting while NQF 3539e assesses patients 65 years of age and older in an inpatient setting. The intent of NQF 3539e is to measure the proportion of older adults with a new antipsychotic prescription in an inpatient setting, however, NQF 3313’s intent is to ensure proper follow-up in an outpatient setting after a new antipsychotic medication prescribed in an outpatient setting.

**[Response Ends]**

**5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.**

*Provide analyses when possible.*

**[Response Begins]**

Not applicable. There are no other NQF-endorsed measures that conceptually address the same measure focus and same target population.

**[Response Ends]**