NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3539e

Corresponding Measures:

De.2. Measure Title: Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting

Co.1.1. Measure Steward: Centers for Medicare & Medicare Services

De.3. Brief Description of Measure: Proportion of inpatient hospitalizations for patients 65 years of age and older who receive an order for antipsychotic medication therapy.

1b.1. Developer Rationale: Clinical guidelines recommend against using antipsychotics as a standard first line of treatment for patients experiencing delirium or behavioral and psychological symptoms of dementia unless they present a threat to themselves or others (AGS 2019, AGS 2015b, NICE 2016, Reus 2016). Antipsychotics are often used off-label as a method of treating patients in an acute confusional state despite conflicting evidence regarding the effectiveness of antipsychotics in treating these disorders (Neufeld 2016, Thom, Mock, and Teslyar et al., 2017, Cascella et al., 2019). Antipsychotic use puts patients with dementia at a greater risk of stroke, cognitive decline, and mortality (AGS 2019). The benefits of this measure lie in the potential to reduce inappropriate use of antipsychotics in inpatient hospital settings and the unnecessary continuation of the intervention post-discharge, resulting in improved patient outcomes (reduced morbidity and mortality) for older adults. Measuring the use of antipsychotics among hospitalized older adult patients could help shift the focus to determining underlying causes of this behavior (such as medication interactions, infection, or sleep disturbances) and adjusting treatment accordingly.

American Geriatrics Society Beers Criteria Update Expert Panel. (2019). American Geriatrics Society 2019 Updated AGS Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc, 67(4):674-694. doi: 10.1111/jgs.15767

AGS Expert Panel on Postoperative Delirium in Older Adults. "American Geriatrics Society abstracted clinical practice guideline for postoperative delirium in older adults." J Am Geriatr Soc, 63(1), 2015b, pp 142-50. doi: 10.1111/jgs.13281.

Cascella M., Fiore, M., Leone, S., Carbone, D., & Di Napoli, R. (2019). Current controversies and future perspectives on treatment of intensive care unit delirium in adults. World journal of critical care medicine, 8(3), 18–27. doi:10.5492/wjccm.v8.i3.18

Neufeld, K.J., Yue, J., Robinson, T.N., et al. (2016). Antipsychotic Medication for Prevention and Treatment of Delirium in Hospitalized Adults: A Systematic Review and Meta-Analysis. J Am Geriatr Soc, 64(4), 705-714.

NICE (National Institute for Health and Clinical Excellence) Dementia: Supporting people with dementia and their careers in health and social care. 2016 (Issued November 2006, Modified September 2016).

Reus, V.I., Fochtmann, L.J., Eyler, A.E., et al. (2016). The American Psychiatric Association Practice Guideline on the Use of Antipsychotics to Treat Agitation or Psychosis in Patients with Dementia. Am J Psychiatry, 173(5), 543-546.

Thom, R.P., Mock, C.K., Teslyar P. (2017). Delirium in hospitalized patients: risks and benefits of antipsychotics. Cleveland Clinic Journal of Medicine, 84 (8), 616-622.

S.4. Numerator Statement: Inpatient hospitalizations for patients who received an order for an antipsychotic medication during the inpatient encounter.

S.6. Denominator Statement: Non-psychiatric inpatient hospitalizations for patients who are 65 and older.

S.8. Denominator Exclusions: Inpatient hospitalizations for patients with a diagnosis of schizophrenia, Tourette´s syndrome, bipolar disorder, Huntington´s disease during the encounter.

Inpatient hospitalizations for patients who were taking antipsychotics prior to admission.

De.1. Measure Type: Process

S.17. Data Source: Electronic Health Records

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not applicable; this measure is not a paired or grouped measure.

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

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

The developer provides the following evidence for this measure:

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

Evidence Summary

- This measure was submitted to the NQF in Fall 2017 as 3315e and reviewed by the Behavioral Health Standing Committee. The Committee encouraged the developer to adjust the measure based on their feedback and bring it back for evaluation in a future endorsement review cycle. Since then, the specifications have been updated to include another denominator exclusion: Inpatient hospitalizations for patients who were taking antipsychotics prior to admission. Other exclusions (e.g., patients taking

antipsychotics who had a depression diagnosis) were also explored based on the Committee's discussions.)

- The logic model provided describes the steps between the healthcare structures and processes and patient's health outcome(s).
- Clinical Practice Guideline recommendations on the use of Antipsychotics (2019 submission):
    - American Geriatrics Society 2019 Updated AGS Beers Criteria for Potentially Inappropriate Medication Use in Older Adults
        - Recommendation to avoid antipsychotics, except in schizophrenia or bipolar disorder, or for short-term use as antiemetic during chemotherapy. **Moderate grade assigned to the evidence and Strong grade assigned to the recommendation.**
    - American Geriatrics Society 2019 American Geriatrics Society Beers Criteria® for Potentially Inappropriate Medication Use in Older Adults Due to Drug-Disease or Drug-Syndrome Interactions That May Exacerbate the Disease or Syndrome
        - Avoid in older adults with or at high risk of delirium because of potential of inducing or worsening delirium. **Moderate grade assigned to the evidence and Strong grade assigned to the recommendation.**
        - Avoid for behavioral problems of dementia and/or delirium unless nonpharmacological options have failed or are not possible and the older adult is threatening substantial harm to self or others. **Moderate grade assigned to the evidence and Strong grade assigned to the recommendation**
- The guidelines state that use of antipsychotics in these cases increases risk of cerebrovascular accident and greater rate of cognitive decline and mortality in persons with dementia.
- In alignment with the evidence provided above, the measure has exclusions for schizophrenia and bipolar disorder. Antipsychotics primarily used as an antiemetic were removed from the medication list. The measure excludes patients who were identified as threatening harm to self or others.

- Clinical Practice Guideline recommendations on the use of Antipsychotics (2017 submission):
    - American Geriatrics Society Guideline 2015 Beers Criteria for Potentially Inappropriate Medication Use in Older Adults recommendation to avoid antipsychotics (except for schizophrenia and bipolar disorder, or as short-term use as antiemetic during chemotherapy). **Moderate grade assigned to the evidence and Strong grade assigned to the recommendation** (e.g. the benefits clearly outweigh harms).
    - American Psychiatric Association Guideline on the Use of Antipsychotics to Treat Agitation or Psychosis in Patients with Dementia (2016). **Moderate grade assigned to the evidence** and **Recommendation grade assigned to the recommendation** (e.g. indicates confidence that the benefits clearly outweigh harms).

Exception to evidence
N/A

Questions for the Committee:

- Do the measure updates (e.g., excluding hospitalizations for patients who were taking antipsychotics prior to admission) and the updated evidence provided strengthen the evidence for this measure?
- How strong is the evidence for this relationship between this measure as specified and patient outcomes?

Guidance from the Evidence Algorithm

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

The highest possible rating is high.

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

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

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

The developer provides rationale for this measure that include the potential benefit to reduce inappropriate use of antipsychotics in inpatient hospital settings and the unnecessary continuation of the intervention post-discharge, resulting in improved patient outcomes for older adults after discharge.

The developer provides performance results at the facility level from two health systems, which provided data from 10 hospitals, and one critical access hospital. The data were derived from three test sites using two different EHRs and representing 137,817 hospital encounters.

Overall summary statistics from all three sites (October 2014 - September 2015)

Mean	Standard Deviation	Min	10th Percentile	Interquartile Range	90th Percentile	Max
17.8%	5.2%	5.5%	8.4%	5.9%	20.4%	22.8%

The developer also provides additional testing for test site 1 from January-December 2018. Mean performance at this site during this time was 30.3% (95% CI, (29.5, 31.1)).

The developer cites additional research on the use of antipsychotics during inpatient hospital visits to support opportunity for improvement:

- A retrospective cohort study of roughly 18,000 adult non-psychiatric hospital admissions over a year found antipsychotic exposure in 9 percent of visits (Herzig 2016a).
- A retrospective cohort study of 2,700,000 adult non-psychiatric hospital admissions over a year found antipsychotic exposure in 6 percent of visits (Marshall 2016).

Disparities

The developer notes that the research on disparities in the use of antipsychotics is limited. They provide analysis of antipsychotic ordering rates on insurance coverage, gender and race (using 2014-2015 data):

- Medicare and Medicaid coverage had the highest rate of antipsychotic ordering.
- Statistically significant differences exist in antipsychotic orders for gender (males higher than females) and race (blacks higher than whites).

The developer examined antipsychotic use disparities using 2018 data. The disparities identified in the prior round of testing persisted across the denominator exclusion conditions tested.

Questions for the Committee:

- Does the data and evidence presented support a gap in care that warrants a national performance measure?
- Does that data support that there is a reasonable benchmark for this measure where differences in performance represent quality differences?

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

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

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

- yes,
- The evidence does suggest that giving these medications wot old adults can be harmful and should only bne used when the patient might be in danger of hurting themselves or others
- Evidence including practice guidelines strongly support avoiding use of antipsychotics in the geriatric population
- tangential
- Evidence based clinical guidelines recommend against using antipsychotics in patients with delirium or sx of dementia. this measure is based on using AGS Beers criteria, which are consensus criteria. This could lead to misuse, recommendation may not apply to individual patients.
- Evidence is supported by a clear and direct causal pathway

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

- yes
- Seems there is a performance gap. For the 65 - 71 age group 21.1 - 28.8% of patients receive these medications. For the over 85 age group 41.0 - 50.5% of patients receive these medications.
- The measure developer demonstrated there is a gap in care and that there are disparities among groups by age, payor, sex, and race.
- Moderate
- Performance gap exists between Medicare/Medicaid and private insurance; males vs females; increased age; and race.
- Yes, clear gap. Disparities demonstrated (gender) but not other characteristics (e.g., racial, SES).

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: NQF Staff

NQF Staff Review

Link to reliability testing results evaluation summary

Link to validity testing results evaluation summary

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- NQF staff is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure, specifically, the accuracy/capture of the exclusion data elements?
- Do the unstructured fields ('Threat to Themselves' and 'Threat to Others') or the non-authoritative source fields ('Bipolar disorder', 'Huntington's disease', 'Schizophrenia', 'Tourette's',) impact the validity of the measure?

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

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

NQF Staff Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3539e

Measure Title: Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting

Type of measure:

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☐ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

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

Level of Analysis:

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

2. **Briefly summarize any concerns about the measure specifications.**
   - Based on the eCQM technical evaluation, the measure specifications follows eCQM industry specs. The measure specifications are fully represented and are not hindered by any limitations in the eCQM industry specs.

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   - Split-half correlation testing was performed to assess the reliability of the performance measure scores. The developer used electronically extracted EHR data from 11 hospitals for reliability testing.
   - For the 2019 submission, the developer used electronically extracted EHR data from 9 hospitals at Test Site 1 to examine the reliability of the measure performance rate. The split-half correlation examines reliability scores for 4 difference exclusion scenarios based on 2018 data.

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

   Submission document: Testing attachment, section 2a2.3

- Results based on 2013-2015 data: The reliability coefficient across 11 hospitals for the antipsychotic measure was 0.98 (with a 95 percent confidence interval, (0.96, 0.99) for all encounters, 65 years of age and older.
- Based on 2018 data, the reliability coefficient across the 9 hospitals for the measure, as originally specified, was 0.95 (with a 95 percent confidence interval, (0.89, 0.99). The reliability coefficient for the three additional conditions, antipsychotics prior to admission (exclusion included in measure specs), antipsychotics for treatment resistant depression, and antipsychotics prior to admission and/or antipsychotics with a treatment-resistant depression diagnosis were 0.95 (0.89, 0.99), 0.95 (0.88, 0.99), and 0.95 (0.88, 0.99), respectively.
- The developer included simulated data set results demonstrating unit testing covering 100% of the measure logic (required for eCQMs).

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)

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

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

   ☒ **Not applicable** (data element testing was not performed)

   The developer does provide data element validity testing, which is described in the validity section.

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

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

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

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

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

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

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    - Testing provided shows the frequency of exclusions and performance rates with and without exclusions (Table 6a). The developer also used experts to help decide which exclusions to add based on empirical analysis and clinical expertise.

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

    Submission document: Testing attachment, section 2b4.

- No Concerns. Developer provides performance differences by hospital, age, sex, race, ethnicity, payer, and ICU exposure. The developer also looks at differences in measure performance based on these characteristics (excluding ICU exposure) and different exclusion scenarios.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Testing attachment, section 2b5.
    - N/A
15. **Please describe any concerns you have regarding missing data.**
Submission document: Testing attachment, section 2b6.
    - There are feasibility concerns noted for some of the exclusion data elements.
16. **Risk Adjustment**

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

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

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d. **Risk adjustment summary:**

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

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

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

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

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

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

- N/A

VALIDITY: TESTING

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

☒ **Face validity**

☐ **Empirical validity testing of the measure score**

☐ **N/A (score-level testing not conducted)**

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

Submission document: Testing attachment, section 2b2.2

Data element validity

- The developer randomly selected a sample of encounters from each test site's EHR extract and manually abstracted data for those encounters in order to assess the chance-adjusted agreement between the two sources. Manual abstraction was done by trained medical record abstractors. A total of 158 encounters were abstracted across test sites.

- To test the additional exclusion data element for the 2019 submission the developer randomly selected a sample of encounters from Test Site's 1 EHR extract and manually abstracted data for those encounters in order to assess the chance-adjusted agreement between the two sources. Manual abstraction was done by trained medical record abstractors. A total of 200 encounters were abstracted.

Face validity

- The developer tested face validity via interviews and a brief web survey from clinicians, information technology professionals, subject matter experts, and members of the expert workgroup (n=8 respondents).
- To gather input on the face validity of the additional exclusion criteria the developer conducted in-depth interviews with representatives from two hospital systems and two members of the Antipsychotics Measure Development EWG.

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

Submission document: Testing attachment, section 2b2.3

Data element validity

- The overall sample of 158 encounters showed 98 percent agreement or higher for all data elements and data element combinations assessed. Agreement was perfect for two of the exclusionary data elements (Tourette's and Huntington's) and the numerator data element (antipsychotic prescription) and almost perfect for the remaining data elements. Kappa values ranged from a low of 0.39 for the numerator exclusion ("threat of harm") to a high of 1.0 for the numerator (medication orders). The numerator exclusion sensitivity is reflective of the inconsistent documentation of the numerator exclusion ("threat of harm") in the EHR.
- Regarding the validity of the new exclusion criterion, antipsychotics prior to admission, there were 36 encounters with an antipsychotic prior to admission in the manually abstracted data and 163 without. In the electronic extract, 180 encounters matched for an agreement rate of 90.5 percent.
- Measure level validity scores with or without exclusions are also provided. Results demonstrate that adding the exclusion strengthens the measure's validity.

Face validity

- Six out of eight respondents reported that hospitals would score well on the measure if they consistently documented "threat of harm" and denominator exclusions.
- Based on input from experts and clinicians, antipsychotics prior to admission is an appropriate denominator exclusion. Those interviewed did not agree with having antipsychotics for the treatment of depression as a denominator exclusion because antipsychotics are not commonly used to treat that condition and the number of patients impacted would be very small.

Based on the eCQM technical review, the following data elements were assessed as having feasibility issues in the accuracy domain.

- Symptom: Threat to themselves or others
- Diagnosis: Bipolar Disorder
- Diagnosis: Huntington's Disease
- Diagnosis: Schizophrenia
- Diagnosis: Tourette's Syndrome

Developer response and plan related to data accuracy:
Data accuracy. All data elements were deemed accurate in the EHRs at the two largest test sites (Test Site 1 and Test Site 2). Test Site 3 reported accuracy issues with the following data elements – 'Bipolar disorder', 'Huntington's disease', 'Schizophrenia', 'Tourette's', 'Threat to Themselves', and 'Threat to Others.' The latter two data elements ('Threat to Themselves' and 'Threat to Others') scored zero on

accuracy because they are not collected in the sites EHR in a structured field. The other data elements, the denominator exclusion conditions, were reported to be accurate but not from the most authoritative source (a score of '2' on the prior version of the feasibility scorecard). This means rather than being based on a test result, the conditions are typically self-reported by patients. Since the current version of the feasibility score has a 0, 1 scoring system, we reduced any scores of less than 3 on the older version of the scorecard to a 0 on the newer version.

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (score-level testing was not performed)

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

    NOTE that data element validation from the literature is acceptable.

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

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (data element testing was not performed)

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

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

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

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

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

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

    - How impactful are the eCQM feasibility issues regarding the accuracy of the following data elements: threat to themselves or others, Bipolar Disorder, Huntington's Disease, Schizophrenia, Tourette's Syndrome on the measure's validity?
    - How concerning is the poor element level validity testing result for "threat of harm"?
    - Minor concern that the additional exclusion element, antipsychotics prior to admission, was only tested in one EHR system.

ADDITIONAL RECOMMENDATIONS

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

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

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

- the developers revised the specification in response to the committee's concerns and excluded individuals with prior antipsychotic use
- This measure seems very reliable due to the specifics of the measure
- The specifications are clearly defined.
- not reliable does not differentiate those with treatment exceptions
- No concerns.
- Split half testing reliable. No need to vote.

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

- yes
- No
- Reliability testing demonstrated adequate reliability. Data suggested documentation of danger to self or others was not consistent among the healthcare organizations.
- no
- No concerns.
- No

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

- yes
- No
- Face validity utilized 8 respondents; data validity was demonstrated using kappa agreement which was 0.38--fair correlation
- no
- face validity used, no concerns.
- Face validity is reasonable.

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

- no concerns
- I did not see any threats to validity
- no significant threats identified
- lots of exclusions
- No concerns.
- Seems to have been adequately assessed.

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

- no concerns
- The additional exclusions provided by the developer help to reduce any threats to the validity of the measure
- Exclusions in the re-specified measure are appropriate
- none
- question about age, over 65 is very broad and doesn't account for functionality of individual.
- Looks ok to me. Using the exclusions improves validity.

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.

- The developer states that all data elements are in defined fields in EHRs and generated during the provision of care.
- This measure is an eCQM. There are no fees required to use the measure.
- Results from the feasibility scorecard are provided.

eCQM technical review

- The following data elements were assessed as having feasibility issues in the workflow domain in one of the EHR systems tested, indicating that the data element is not routinely generated and used during care delivery:
    - Symptom: Threat to themselves or others
    - Diagnosis: Bipolar Disorder
    - Diagnosis: Huntington's Disease
    - Diagnosis: Schizophrenia
    - Diagnosis: Tourette's Syndrome
- The following data element was assessed as having feasibility issues in the availability/standards domain(s) in one of the EHR systems tested, indicating that the data element may not be available electronically or have a credible near term path to electronic collection:
    - Symptom: Threat to themselves or others
- All value sets used in measure submission are accessible via the VSAC.

- The developer included simulated data set results demonstrating unit testing covering 100% of the measure logic.

Developer response and plan related to data availability, data standards, and workflow:

Data availability. Based on our feasibility assessment, all data elements were deemed available in the EHRs at the two largest test sites (Test Site 1 and Test Site 2). The third test site, a critical access hospital (CAH) in rural Pennsylvania, had challenges with the availability of the numerator exclusion data elements – 'threat to themselves' and 'threat to others'. Test Site 3 reported that as a CAH, they do not treat patients that are a threat to themselves or to others. Rather, patients exhibiting these behaviors are typically transferred to another facility for care. Test Site 3 documents behavioral information in free text fields. If this measure were to be implemented, Test Site 3 indicated that they would update their EHR so that 'threat to themselves' and 'threat to others' are documented in structured fields to allow for data extraction and reporting.

Data Standards. All data elements were coded using nationally accepted terminologies at Test Site 1. Test Site 3 scored 0 on data standards for 'threat to themselves' and 'threat to others' because neither are available in a structured field in its EHR.

Workflow: At Test Site 1, all data elements required for the measure calculation are collected during standard care. Both Test Sites 2 and 3 would require changes in their workflows to regularly capture the following data elements - Bipolar disorder, Huntington's disease, Schizophrenia, Tourette's, "Threat to Themselves, and Threat to Others. In order to incorporate these data elements into the Test Sites' regular workflow, there would need to be modifications to the EHR systems and staff training. Both test sites indicated that these changes would be made if reporting on the antipsychotic measure was required.

Questions for the Committee:

- Are the required data elements, specifically the exclusion elements listed above, routinely generated and documented?
- Are the feasibility scorecard results and the developer's rationale for the low-scoring data elements acceptable?
- Can this measure be widely implemented and provide accurate results?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- yes
- All data elements are available via the EHR and/or ePrescribing system
- Some of the exclusionary criteria (e.g. threat to self or others) may be hard to find in the record; measure would necessitate organizations modify their EMRs
- risk to self not identified
- Definition of "threat of harm" to self or others may not be clear.

- Feasible--may require workflow changes, but doable.

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details

- The developer notes that CMS is considering implementation plans for this measure.
- The measure has been submitted through the Measures Under Consideration process for the CMS Hospital Inpatient Quality Reporting Program and the Medicare and Medicaid Promoting Interoperability Program for Eligible Hospitals and Critical Access Hospitals.

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

N/A – the measure has not been implemented.

Additional Feedback:

N/A

Questions for the Committee:

- Does the measure have demonstrated potential to help inform healthcare decisions to reduce the inappropriate use of antipsychotics in the inpatient hospital setting?

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

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

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

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

Improvement results

- The developer states that this measure could be used to improve the quality and safety of care for hospitalized older adults and encourage more judicious prescribing of antipsychotics for hospitalized patients. These actions may additionally result in fewer prescriptions after discharge reducing morbidity and mortality associated with long-term use of these medications.

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

Unexpected findings (positive or negative) during implementation

- The measure has not yet been implemented. Potential benefits (beyond evidence already presented) include thoughtful prescribing of antipsychotics in the inpatient setting, fewer continued prescriptions after discharge, use of delirium assessment and monitoring tools, improved detection of patient behaviors that could escalate, and use of nonpharmacologic interventions.

Potential harms

- Potential unintended consequences include increased use of alternative harmful medications such as benzodiazepines for delirium or behavioral and psychological symptoms of dementia and hesitancy using antipsychotics when warranted.

Additional Feedback:

- N/A

Questions for the Committee:

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

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

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

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided?4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- not used yet
- Yes that data was provided to the test hospitals
- New measure; meant to distinguish performance among hospitals.
- yes

- Feedback has been given.
- No data but has potential.

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

- usable. unintended consequences would be to deter prescription of antipsychotic medication when needed.
- Prescribers might err on the side of meeting the measure and not adequately weigh the benefit to the individual patient in front of them who may actually need these medications (real life situation)
- Performance gap demonstrated and it is hoped that the measure would help close it. Unintended consequences were identified including increase use of restraints or use of other medications more inappropriate than antipsychotics.
- usable
- Could be unintended consequences, antipsychotics may not be used when they are appropriate. Other medications, such as benzos may be used with unintended consequences.
- Looks promising.

Criterion 5: Related and Competing Measures

Related or competing measures

Related measures
2111 : Antipsychotic Use in Persons with Dementia
2993 : Potentially Harmful Drug-Disease Interactions in the Elderly

Harmonization
Developer states that the measures are harmonized to the extent possible. The submitted measure is the only "inappropriate use" measure of antipsychotic medications in the inpatient hospital setting.

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

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

- not that I know of
- There are two that the developer identified - 2111 : Antipsychotic Use in Persons with Dementia and 2993 : Potentially Harmful Drug-Disease Interactions in the Elderly
- There are related measures with different exclusions. However none focus on the inpatient population.
- none known
- There are related measures, with efforts for harmonization.
- What are the specs on 2111? This is a fairly focused measure (as opposed to 2993).

Public and Member Comments

No comments or member support/non-support choices have been submitted as of 11/25/2019.

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

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

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

2._Hospital-MDM_NQF_Form_Evidence_v7.1_AP.docx

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

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

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Not applicable

Measure Title: Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Not applicable

Date of Submission: TBD

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

Outcome

☐ Outcome: Click here to name the health outcome

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

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

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

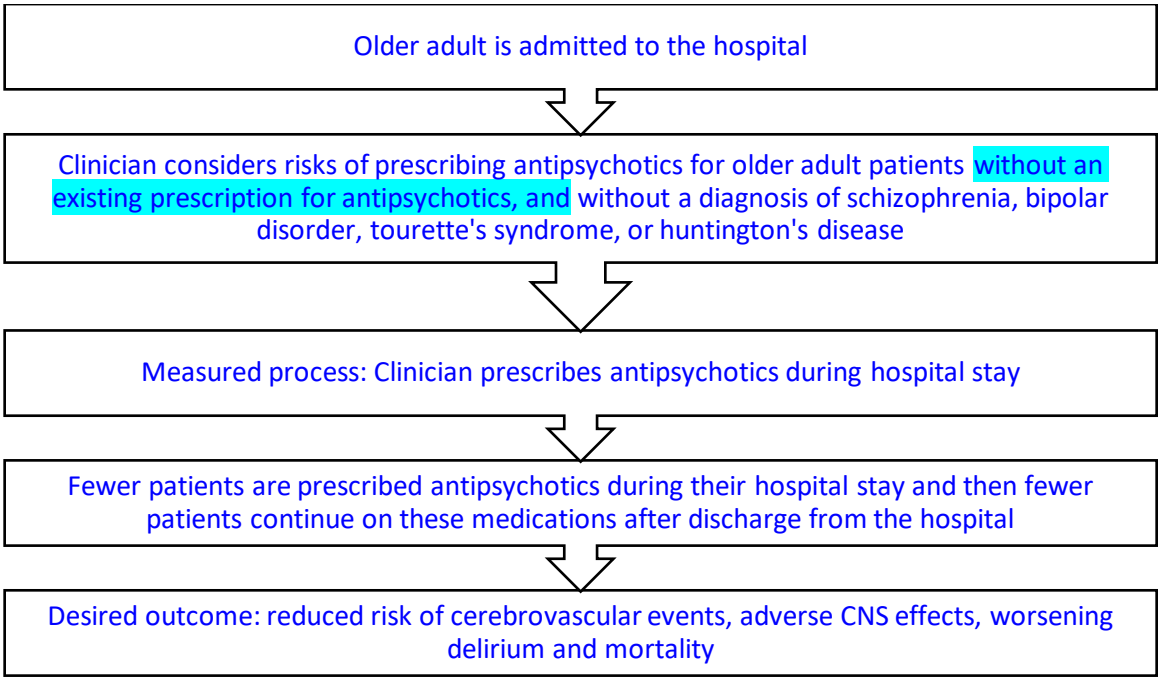
☒ Process: Prescribing of potentially inappropriate medications for older adults

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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



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

Not applicable.

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

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

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Table 1. American Geriatrics Society Guideline

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**Title: American Geriatrics Society 2019 Updated AGS Beers Criteria for Potentially Inappropriate Medication Use in Older Adults** **Author: 2019 American Geriatrics Society Beers Criteria Update Expert Panel.** **Date: 2019.** **Citation: The 2019 American Geriatrics Society Beers Criteria® Update Expert Panel. 2019. American Geriatrics Society 2019 Updated AGS Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. Journal of the American Geriatrics Society, 63(11): 2227-2246.** **URL: http://geriatricscareonline.org/ProductAbstract/american-geriatrics-society-updated-beers-criteria-for-potentially-inappropriate-medication-use-in-older-adults/CL001**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	"The primary target audience for the AGS Beers Criteria® is practicing clinicians. The criteria are intended for use in adults 65 years and older in all ambulatory, acute, and institutionalized settings of care, except for the hospice and palliative care settings. Consumers, researchers, pharmacy benefits managers, regulators, and policymakers also widely use the AGS Beers Criteria®. The intention of the AGS Beers Criteria® is to improve medication selection; educate clinicians and patients; reduce adverse drug events; and serve as a tool for evaluating quality of care, cost, and patterns of drug use of older adults." **Table 2.** 2019 American Geriatrics Society Beers Criteria® for Potentially Inappropriate Medication Use in Older Adults **Organ System, Therapeutic Category, Drug(s):** Antipsychotics, first (conventional) and second (atypical) generation **Rationale:** "Increased risk of cerebrovascular accident (stroke) and greater rate of cognitive decline and mortality in persons with dementia Avoid antipsychotics for behavioral problems of dementia or delirium unless nonpharmacological options (eg, behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others" **Recommendation:**

	"Avoid, except in schizophrenia or bipolar disorder, or for short-term use as antiemetic during chemotherapy" **Table 3.** 2019 American Geriatrics Society Beers Criteria® for Potentially Inappropriate Medication Use in Older Adults Due to Drug-Disease or Drug-Syndrome Interactions That May Exacerbate the Disease or Syndrome **Disease or Syndrome:** Delirium **Drug(s):** Antipsychotics* **Rationale:** "Avoid in older adults with or at high risk of delirium because of potential of inducing or worsening delirium Avoid antipsychotics for behavioral problems of dementia and/or delirium unless nonpharmacological options (eg, behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others. Antipsychotics are associated with greater risk of cerebrovascular accident (stroke) and mortality in persons with dementia." **Recommendation:** Avoid *May be required to treat concurrent schizophrenia, bipolar disorder, and other selected mental health conditions but should be prescribed in the lowest effective dose and shortest possible duration. **Table 3.** 2019 American Geriatrics Society Beers Criteria® for Potentially Inappropriate Medication Use in Older Adults Due to Drug-Disease or Drug-Syndrome Interactions That May Exacerbate the Disease or Syndrome **Disease or Syndrome:** Dementia or cognitive impairment **Drug(s):** Antipsychotics, chronic and as-needed use * **Rationale:** "Avoid because of adverse CNS effects Avoid antipsychotics for behavioral problems of dementia and/or delirium unless nonpharmacological options (eg, behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others. Antipsychotics are associated with greater risk of cerebrovascular accident (stroke) and mortality in persons with dementia." **Recommendation:** Avoid

	*May be required to treat concurrent schizophrenia, bipolar disorder, and other selected mental health conditions but should be prescribed in the lowest effective dose and shortest possible duration.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**Table 2.** Quality of evidence: moderate **Table 3. Delirium** Quality of evidence: moderate **Table 3. Dementia or cognitive impairment** Quality of evidence: moderate "Evidence…obtained from RCTs with important limitations…. In addition, evidence from well-designed controlled trials without randomization, well-designed cohort or case-control analytic studies, and multiple time series with or without intervention are in this category. Moderatequality evidence also means that further research will probably have an important effect on our confidence in the estimate of effect and may change the estimate."
Provide all other grades and definitions from the evidence grading system	Quality of evidence: high "Evidence…obtained from 1 or more well-designed and well-executed randomized, controlled trials (RCTs) that yield consistent and directly applicable results. This also means that further research is very unlikely to change our confidence in the estimate of effect." Quality of evidence: low "Evidence obtained from observational studies would typically be rated as low quality because of the risk for bias. Low-quality evidence means that further research is very likely to have an important effect on our confidence in the estimate of effect and will probably change the estimate. However, the quality of evidence may be rated as moderate or even high, depending on circumstances under which evidence is obtained from observational studies."
Grade assigned to the **recommendation** with definition of the grade	**Table 2.** Strength of recommendation: strong **Table 3. Delirium** Strength of recommendation: strong **Table 3. Dementia or cognitive impairment**

	Strength of recommendation: strong "Harms, adverse events, and risks clearly outweigh benefits."
Provide all other grades and definitions from the recommendation grading system	Strength of recommendation: weak "Harms, adverse events, and risks may not outweigh benefits."
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	The quantity and quality of evidence for the 2019 Beers Criteria update are similar to what was included in the 2015 Beers Criteria. AGS did not provide detailed information on each supporting study for the 2019 Beers Criteria update, limiting our ability to characterize the evidence in detail.
Estimates of benefit and consistency across studies	The benefit and consistency across studies for the 2019 Beers Criteria update are similar to what was included in the 2015 Beers Criteria. AGS did not provide detailed information on each supporting study for the 2019 Beers Criteria update, limiting our ability to characterize the evidence in detail.
What harms were identified?	The identified harms for the 2019 Beers Criteria update are similar to what was included in the 2015 Beers Criteria. AGS did not provide detailed information on each supporting study for the 2019 Beers Criteria update, limiting our ability to characterize the evidence in detail.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	No recent studies change the conclusions of the 2019 Beers Criteria.

Table 2. American Geriatrics Society Guideline

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**Title: American Geriatrics Society 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults.** **Author: American Geriatrics Society 2015 Beers Criteria Update Expert Panel.** **Date: 2015.** **Citation: American Geriatrics Society 2015 Beers Criteria Update Expert Panel. 2015. American Geriatrics Society 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. Journal of the American Geriatrics Society, 63(11): 2227-2246.** **URL: http://geriatricscareonline.org/ProductAbstract/american-geriatrics-society-updated-beers-criteria-for-potentially-inappropriate-medication-use-in-older-adults/CL001**

Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	"The primary target audience for the AGS Beers Criteria is practicing clinicians. The criteria are intended for use in all ambulatory, acute, and institutionalized settings of care for populations aged 65 and older in the United States, with the exception of hospice and palliative care. Consumers, researchers, pharmacy benefits managers, regulators, and policymakers also widely use the AGS Beers Criteria. The intentions of the criteria are to: improve medication selection; educate clinicians and patients; reduce adverse drug events; and serve as a tool for evaluating quality of care, cost, and patterns of drug use of older adults." **Table 2** 2015 American Geriatrics Society Beers Criteria for Potentially Inappropriate Medication Use in Older Adults **Organ System, Therapeutic Category, Drugs:** Antipsychotics, first- (conventional) and second- (atypical) generation **Rationale:** Increased risk of cerebrovascular accident (stroke) and greater rate of cognitive decline and mortality in persons with dementia Avoid antipsychotics for behavioral problems of dementia or delirium unless nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others **Recommendation:** Avoid, except for schizophrenia, bipolar disorder, or short-term use as antiemetic during chemotherapy **Table 3** 2015 American Geriatrics Society Beers Criteria for Potentially Inappropriate Medication Use in Older Adults Due to Drug-Disease or Drug-Syndrome Interactions That May Exacerbate the Disease or Syndrome **Disease or Syndrome:** Delirium **Drug(s):** Antipsychotics **Rationale:** Avoid in older adults with or at high risk of delirium because of the potential of inducing or worsening delirium Avoid antipsychotics for behavioral problems of dementia or delirium unless nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others Antipsychotics are associated with greater risk of cerebrovascular accident (stroke) and mortality in persons with dementia **Recommendation:** Avoid

	Table 3 2015 American Geriatrics Society Beers Criteria for Potentially Inappropriate Medication Use in Older Adults Due to Drug-Disease or Drug-Syndrome Interactions That May Exacerbate the Disease or Syndrome **Disease or Syndrome:** Dementia or cognitive impairment **Drug(s):** Antipsychotics, chronic and as-needed use **Rationale:** Avoid because of adverse CNS effects Avoid antipsychotics for behavioral problems of dementia or delirium unless nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others. Antipsychotics are associated with greater risk of cerebrovascular accident (stroke) and mortality in persons with dementia **Recommendation:** Avoid
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Moderate: Evidence is sufficient to determine risks of adverse outcomes, but the number, quality, size, or consistency of included studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes (≥1 higher-quality trial with >100 participants; ≥2 higher-quality trials with some inconsistency; ≥2 consistent, lower-quality trials; or multiple, consistent observational studies with no significant methodological flaws showing at least moderate effects) limits the strength of the evidence
Provide all other grades and definitions from the evidence grading system	High: Evidence includes consistent results from well designed, well-conducted studies in representative populations that directly assess effects on health outcomes (≥2 consistent, higher-quality randomized controlled trials or multiple, consistent observational studies with no significant methodological flaws showing large effects) Low: Evidence is insufficient to assess harms or risks in health outcomes because of limited number or power of studies, large and unexplained inconsistency between higher-quality studies, important flaws in study design or conduct, gaps in the chain of evidence, or lack of information on important health outcomes
Grade assigned to the **recommendation** with definition of the grade	Strong: Benefits clearly outweigh harms, adverse events, and risks, or harms, adverse events, and risks clearly outweigh benefits
Provide all other grades and definitions from the recommendation grading system	Weak: Benefits may not outweigh harms, adverse events, and risks Insufficient: Evidence inadequate to determine net harms, adverse events, and risks
Body of evidence:	The Beers Criteria were first published in 1991. Since that

• Quantity – how many studies? • Quality – what type of studies?	time the criteria have been regularly updated based off of the existing criteria and any new evidence published since the last update. The American Geriatrics Society forms an expert panel to update the Beers Criteria every few years. The panel works from the previous evidence review and then reviews any new evidence published since that last review to update the recommendations in the Beers Criteria. The 2015 review by the AGS 2015 Beers Criteria Update Expert Panel included review of 60 systematic reviews and meta analyses, 49 randomized control trials (RCTs) and 233 observational studies and other types of publications. Overall, the quality of the evidence is good. In addition to conducting a systematic review of the evidence, the AGS 2015 Beers Criteria Update Expert Panel also used technical experts and a public comment period for additional validity. **Table 2, antipsychotics:** Evidence for the recommendation to avoid antipsychotics in older adults was rated as moderate quality. It includes 2 randomized control studies, 3 systematic reviews, 2 cohort studies and 1 observational study. **Table 3, delirium:** Evidence for the recommendation to avoid certain medications (including antipsychotics) for individuals with delirium was rated as moderate quality. It includes 2 systematic reviews, 1 randomized controlled study, 8 cohort studies, 1 observational study and 1 clinical review. **Table 3, dementia or cognitive impairment**: Evidence for the recommendation to avoid certain medications (including antipsychotics) for individuals with dementia was rated as moderate quality. It includes 3 systematic reviews and 2 randomized control studies in addition to 4 cohort studies.
Estimates of benefit and consistency across studies	Recommendations in the Beers criteria are based on studies that explain the rationale for why a medication group is potentially harmful for older adults (Table 2) or for older adults with a certain condition (Table 3). Below is a summary of the number and types of studies supporting the relevant recommendations regarding antipsychotics. Studies consistently found an increased risk of adverse events associated with antipsychotic use. Summaries of each study can be found on the American Geriatrics Society's website: http://www.americangeriatrics.org/. **Table 2, Antipsychotics** Studies that support the recommendation: 2015 Criteria:

	Hwang 2014 – retrospective cohort Langballe 2014 – retrospective cohort From previous criteria: Dore 2009 – observational Maher 2011 – systematic review, meta-analysis Schneider 2005 – systematic review, meta-analysis Schneider 2006a – systematic review, meta-analysis Schneider 2006b – randomized control trial Vigen 2011 – randomized control trial Recommendation: Avoid, except for schizophrenia, bipolar disorder, or short-term use as antiemetic during chemotherapy **Table 3, Delirium** Studies that support the recommendation: From 2015 Criteria: Aparasu 2012 – retrospective cohort Chavant 2011 – retrospective cohort Citrome 2013 – retrospective cohort Hampton 2014 – retrospective cohort Han 2004 – randomized control trial Rigler 2013 – retrospective cohort From previous criteria: Clegg 2011 – systematic review Gaudreau 2005 – prospective cohort Laurila 2008 – observational study Marcantonio 1994a – prospective cohort Moore 1999 – clinical review Ozbolt 2008 – systematic review Rudolph 2008 – retrospective and prospective cohorts Recommendation: Avoid in older adults with or at high risk of delirium because of potential of inducing or worsening delirium Avoid antipsychotics for behavioral problems of dementia and/or delirium unless nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others. Antipsychotics are associated with increased risk of cerebrovascular accident (stroke) and mortality in persons with dementia

	Table 3, Dementia or cognitive impairment Studies that support the recommendation: 2015 Criteria: Chavant 2011 – retrospective cohort Kalicsh Ellet 2014 – retrospective cohort From previous criteria: Rudolph 2008 – retrospective and prospective cohorts Schneider 2005 – systematic review, meta-analysis Schneider 2006a – systematic review, meta-analysis Schneider 2006b – randomized control trial Seitz 2011 – systematic review, meta-analysis Vigen 2011 – randomized control trial Wright 2009 – prospective longitudinal cohort Recommendation: Avoid due to adverse CNS effects Avoid antipsychotics for behavioral problems of dementia and/or delirium unless nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others. Antipsychotics are associated with increased risk of cerebrovascular accident (stroke) and mortality in persons with dementia
What harms were identified?	As part of their review of the evidence, the AGS 2015 Beers Criteria Update Expert Panel identified subgroups of patients who should be exempt from the criteria and for whom listed medications may be appropriate. In addition, a patient could have a condition or comorbidity that would merit the use of a medication on the list, even if the comorbidity is not specifically listed in the criteria. The criteria are designed to assist providers in the prescribing of potentially harmful medications, and should not be taken as strict criteria to avoid use in all patients without weighing the harms and benefits for individual cases. **Table 2, Antipsychotics** "Avoid, except for schizophrenia, bipolar disorder, or short-term use as antiemetic during chemotherapy" The proposed measure has exclusions for schizophrenia and bipolar disorder. Antipsychotics primarily used as an antiemetic were removed from the list of medications for the proposed measure.

	Table 3, Delirium "Avoid antipsychotics for behavioral problems of dementia or delirium unless nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others" The proposed measure has an exclusion for patients who were identified as threatening harm to self or others. **Table 3, Dementia or cognitive impairment** "Avoid antipsychotics for behavioral problems of dementia or delirium unless nonpharmacological options (e.g., behavioral interventions) have failed or are not possible and the older adult is threatening substantial harm to self or others" The proposed measure has an exclusion for patients who were identified as threatening harm to self or others.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Relevant studies have been published since the publication of the guideline, but they do not change these conclusions. Relevant studies include, but are not limited to, the following: Herzig SJ, Rothberg MB, Guess JR, et al. 2016. "Antipsychotic Use in Hospitalized Adults: Rates, Indications, and Predictors." J Am Geriatr Soc 64(2): 299-305. doi: 10.1111/jgs.13943 Marshall J, Herzig SJ, Howell MD, et al. 2016. "Antipsychotic utilization in the intensive care unit and in transitions of care." J Crit Care 33: 119-124. doi: 10.1016/j.jcrc.2015.12.017 Neufeld KJ, Yue J, Robinson TN, et al. 2016. "Antipsychotic Medication for Prevention and Treatment of Delirium in Hospitalized Adults: A Systematic Review and Meta-Analysis." J Am Geriatr Soc 64(4):705-714. doi: 10.1111/jgs.14076

Table 3. American Psychiatric Association Guideline

Source of Systematic Review: • **Title** • **Author** • **Date**	**Title: The American Psychiatric Association Practice Guideline on the Use of Antipsychotics to Treat Agitation or Psychosis in Patients with Dementia** **Author: Reus VI, Fochtmann LJ, Eyler AE, Hilty DM,**

• **Citation, including page number** • **URL**	**Horvitz-Lennon M, Jibson MD, Lopez OL, Mahoney J, Pasic J, Tan ZS, Wills CD, Rhoads R, Yager J.** **Date: 2016** **Citation, including page number: Reus VI, Fochtmann LJ, Eyler AE, et al. 2016. "The American Psychiatric Association Practice Guideline on the Use of Antipsychotics to Treat Agitation or Psychosis in Patients with Dementia." *Am J Psychiatry* 173(5):543-6. doi: 10.1176/appi.ajp.2015.173501.** **URL: http://ajp.psychiatryonline.org/doi/pdf/10.1176/appi.ajp.2015.173501**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	"The goal of this guideline is to improve the care of patients with dementia who are exhibiting agitation or psychosis. More specifically, this guideline focuses on the judicious use of antipsychotic medications when agitation or psychosis occurs in association with dementia and does not review evidence for or focus on other pharmacological interventions. The guideline is intended to apply to individuals with dementia in all settings of care as well as to care delivered by generalist and specialist clinicians. Recommendations regarding treatment with antipsychotic medications are not intended to apply to individuals who are receiving antipsychotic medication for another indication (e.g., chronic psychotic illness) or individuals who are receiving an antipsychotic medication in an urgent context." "Statement 5. APA recommends that nonemergency antipsychotic medication should only be used for the treatment of agitation or psychosis in patients with dementia when symptoms are severe, are dangerous, and/or cause significant distress to the patient. (1B)"
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**"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."
Provide all other grades and definitions from the evidence grading system	"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." "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."

Grade assigned to the **recommendation** with definition of the grade	**"Recommendation"** (denoted by the numeral 1 after the guideline statement) indicates confidence that the benefits of the intervention clearly outweigh harms.
Provide all other grades and definitions from the recommendation grading system	"Suggestion" (denoted by the numeral 2 after the guideline statement) indicates uncertainty (i.e., the balance of benefits and harms is difficult to judge or either the benefits or the harms are unclear).
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Overall, 45 randomized controlled trials and 52 observational studies were included in the guideline.
Estimates of benefit and consistency across studies	"Statements 5, 8, 10, 14, and 15 are based on moderate-strength evidence in individuals with dementia that the benefits of antipsychotic medication are small. In addition, consistent evidence, predominantly from large observational studies, indicates that antipsychotic medications are associated with clinically significant adverse effects, including mortality, among individuals with dementia. The overall strength of evidence for these statements is graded as moderate on the basis of this balance of benefits and harms data and the fact that there were no studies that directly addressed all of the specific elements of each recommendation."
What harms were identified?	The Guideline Writing Group acknowledged that there are some situations where antipsychotic use for patients with dementia may be appropriate: "Expert consensus suggests that use of an antipsychotic medication in individuals with dementia can be appropriate, particularly in individuals with dangerous agitation or psychosis (see "Expert Opinion Survey Data: Results" in Appendix B), and can minimize the risk of violence, reduce patient distress, improve the patient's quality of life, and reduce caregiver burden. However, in clinical trials, the benefits of antipsychotic medications are at best small (Corbett et al. 2014; Kales et al. 2015; see "Review of Supporting Research Evidence" in Appendix A) whether assessed through placebo controlled trials, head-to-head comparison trials, or discontinuation trials."
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Relevant studies have been published since the publication of the guideline, but they do not change these conclusions. Relevant studies include, but are not limited to, the following: Herzig SJ, Rothberg MB, Guess JR, et al. 2016. "Antipsychotic Use in Hospitalized Adults: Rates, Indications, and Predictors." J Am Geriatr Soc 64(2): 299-305. doi: 10.1111/jgs.13943

	Marshall J, Herzig SJ, Howell MD, et al. 2016. "Antipsychotic utilization in the intensive care unit and in transitions of care." J Crit Care 33: 119-124. doi: 10.1016/j.jcrc.2015.12.017 Neufeld KJ, Yue J, Robinson TN, et al. 2016. "Antipsychotic Medication for Prevention and Treatment of Delirium in Hospitalized Adults: A Systematic Review and Meta-Analysis." J Am Geriatr Soc 64(4):705-714. doi: 10.1111/jgs.14076

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

1b. Performance Gap

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

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

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

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

Clinical guidelines recommend against using antipsychotics as a standard first line of treatment for patients experiencing delirium or behavioral and psychological symptoms of dementia unless they present a threat to themselves or others (AGS 2019, AGS 2015b, NICE 2016, Reus 2016). Antipsychotics are often used off-label as a method of treating patients in an acute confusional state despite conflicting evidence regarding the effectiveness of antipsychotics in treating these disorders (Neufeld 2016, Thom, Mock, and Teslyar et al., 2017, Cascella et al., 2019). Antipsychotic use puts patients with dementia at a greater risk of stroke, cognitive decline, and mortality (AGS 2019). The benefits of this measure lie in the potential to reduce inappropriate use of antipsychotics in inpatient hospital settings and the unnecessary continuation of the intervention post-discharge, resulting in improved patient outcomes (reduced morbidity and mortality) for older adults.

Measuring the use of antipsychotics among hospitalized older adult patients could help shift the focus to determining underlying causes of this behavior (such as medication interactions, infection, or sleep disturbances) and adjusting treatment accordingly.

American Geriatrics Society Beers Criteria Update Expert Panel. (2019). American Geriatrics Society 2019 Updated AGS Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc, 67(4):674-694. doi: 10.1111/jgs.15767

AGS Expert Panel on Postoperative Delirium in Older Adults. "American Geriatrics Society abstracted clinical practice guideline for postoperative delirium in older adults." J Am Geriatr Soc, 63(1), 2015b, pp 142-50. doi: 10.1111/jgs.13281.

Cascella M., Fiore, M., Leone, S., Carbone, D., & Di Napoli, R. (2019). Current controversies and future perspectives on treatment of intensive care unit delirium in adults. World journal of critical care medicine, 8(3), 18–27. doi:10.5492/wjccm.v8.i3.18

Neufeld, K.J., Yue, J., Robinson, T.N., et al. (2016). Antipsychotic Medication for Prevention and Treatment of Delirium in Hospitalized Adults: A Systematic Review and Meta-Analysis. J Am Geriatr Soc, 64(4), 705-714.

NICE (National Institute for Health and Clinical Excellence) Dementia: Supporting people with dementia and their careers in health and social care. 2016 (Issued November 2006, Modified September 2016).

Reus, V.I., Fochtmann, L.J., Eyler, A.E., et al. (2016). The American Psychiatric Association Practice Guideline on the Use of Antipsychotics to Treat Agitation or Psychosis in Patients with Dementia. Am J Psychiatry, 173(5), 543-546.

Thom, R.P., Mock, C.K., Teslyar P. (2017). Delirium in hospitalized patients: risks and benefits of antipsychotics. Cleveland Clinic Journal of Medicine, 84 (8), 616-622.

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

The measure was initially tested in two health systems, which provided data from 10 hospitals, and one critical access hospital. These systems, Test Site 1 (in Texas), Test Site 2 (in North Carolina), and Test Site 3 (in Pennsylvania) varied in terms of their EHR product and version and the size of their hospital systems. Test Sites 1 and 2 used different installations of the Cerner EHR product. Test Site 3 used a Meditech EHR product. With respect to size, Test Site 1 had the most beds (n=3,320) and Test Site 3, a critical access hospital, had the least number of beds (n=25). Test Site 1 provided data for nine hospitals in its system. Test Site 2 and Test Site 3 provided data for one hospital each. Across the three test sites, we received data on 137,817 hospital encounters. Test Site 1 contributed the most encounters (n= 99,528) followed by Test Site 2 (n=37,560) and Test Site 3 (n=729). A detailed breakdown of the characteristics of the measured facilities and the patient populations can be found in sections 1.5 and 1.6 of the attached Measure Testing form.

The measure performance, including the denominator, numerator, and the measure rate by hospital, is presented below.

Test site 1:

- Dates of data: October 1, 2013–September 30, 2015
- Denominator: 99,528
- Denominator after exclusions: 92,943
- Numerator: 16,229
- Numerator exclusions: 153
- Numerator after exclusions: 16,076

- Measure rate: 17.3%

Test site 2:

- Dates of data: October 1, 2013–September 30, 2015
- Denominator: 37,560
- Denominator after exclusions: 35,385
- Numerator: 6,984
- Numerator exclusions: 112
- Numerator after exclusions: 6,872
- Measure rate: 19.4%

Test site 3:

- Dates of data: October 1, 2014–September 30, 2015
- Denominator: 729
- Denominator after exclusions: 727
- Numerator: 40
- Numerator exclusions: 0
- Numerator after exclusions: 40
- Measure rate: 5.5%

Overall: (summary statistics based on the October 1, 2014–September 30, 2015, data from Test Sites 1 & 2 and October 1, 2014–September 30, 2015, data from Test Site 3)

- Mean: 17.8%
- Std. Deviation: 5.2%
- Coefficient of variation: 0.342
- Min: 5.5%
- Max: 22.8%
- Interquartile Range: 5.9%
- 10th Percentile: 8.4%
- 25th Percentile: 12.8%
- 50th Percentile: 15.9%
- 75th Percentile: 18.7%
- 90th Percentile: 20.4%

Overall: (summary statistics based on the October 1, 2014–September 30, 2015, data from Test Sites 1 & 2)

- Mean: 16.3%
- Std. Deviation: 4.3%
- Coefficient of variation: 0.265
- Min: 8.4%
- Max: 22.8%
- Interquartile Range: 5.1%
- 10th Percentile: 11.4%
- 25th Percentile: 14.0%

- 50th Percentile: 16.9%

- 75th Percentile: 19.1%

- 90th Percentile: 20.7%

2019 resubmission:

Test site 1:

- Dates of data: January 1, 2018–December 31, 2018

- Denominator: 16,621

- Denominator after exclusions: 14,074

- Numerator: 4271

- Numerator exclusions: 15

- Measure rate: 30.3 (95% CI, (29.5, 31.1))

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

Research on the use of antipsychotics during an inpatient hospital visit is limited, but there are indications of a quality gap. Recent studies have estimated the prevalence of potentially inappropriate antipsychotic use in the inpatient setting.

A retrospective cohort study of roughly 18,000 adult (18 and older) non-psychiatric hospital admissions over a year found antipsychotic exposure in 9 percent of visits. More than half of these were patients who may have been initiated on antipsychotics during those visits. Twenty-six percent of the patients who were initiated on an antipsychotic were then discharged on an antipsychotic. The most common reasons documented for initiating antipsychotics were delirium or probable delirium (Herzig 2016a).

A retrospective cohort study of 2,700,000 adult non-psychiatric hospital admissions over a year found antipsychotic exposure in 6 percent of visits. This rate varied by age, with 4.6 percent of patients age 18–65, 5.2 percent of patients age 65–74, and 8.8 percent of patients age 75 and older being exposed to antipsychotics during their inpatient stay. This study also found that 29 percent of admissions with delirium and 27 percent of admissions with dementia received antipsychotics. This study concluded that there was variation in antipsychotic use between hospitals, which should be explored further (Herzig 2016b). A retrospective cohort study of approximately 39,000 ICU admissions over the course of 7 years found that 8 percent of patients were newly initiated on antipsychotics during the ICU visit and 21 percent of these patients were continued on antipsychotics after discharge (Marshall 2016).

Herzig, S.J., M.B. Rothberg, J.R. Guess, et al. “Antipsychotic Use in Hospitalized Adults: Rates, Indications, and Predictors” J Am Geriatr Soc, 64(2), 2016a, pp 299-305.

Herzig, S.J., M.B. Rothberg, J.R. Guess, et al. “Antipsychotic medication utilization in nonpsychiatric hospitalizations.” J Hosp Med, 11(8), 2016b, pp 543-549.

Marshall, J., S.J. Herzig, M.D. Howell, et al. “Antipsychotic utilization in the intensive care unit and in transitions of care.” J Crit Care, 33, 2016, pp 119-24.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(<u>This is required for maintenance of endorsement</u>. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., “topped out”, disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Data collected during measure testing on older adult patients (65 years and older) found that patients with Medicare and Medicaid coverage had the highest rate of antipsychotic ordering (22.0 percent and 27.9 percent, respectively). Patients with private insurance had the lowest rates at 13.4 percent. Measure testing found statistically significant differences in antipsychotic orders for males compared to females (24.0 and 19.7 percent, respectively). Difference in the rate of antipsychotic ordering by race is significant as well. Across racial groups (black, white, and other), the rate of antipsychotic ordering ranged from 20.9 to 24.4. Hispanic and non-Hispanic patients had similar performance rates, 20.6 and 22.0 respectively. Although the difference is small between the two ethnicity groups and likely not clinically significant, it is statistically significant (p=.022), which is likely due to the large sample size.

2019 resubmission: When testing for the 2019 submission, we considered four denominator exclusion conditions - 1) excluding patients with Tourette´s, Bipolar, Huntington´s, or schizophrenia, 2) excluding patient with Tourette´s, Bipolar, Huntington´s, or schizophrenia and excluding patients taking antipsychotics prior to admission, 3) excluding patients with Tourette´s, Bipolar, Huntington´s, or schizophrenia and excluding patients taking antipsychotics who had a depression diagnosis, and 4) excluding patients with Tourette´s, Bipolar, Huntington´s, or schizophrenia and (excluding patients taking antipsychotics who had a depression diagnosis and/or patients taking antipsychotics prior to admission).

The disparities identified in the prior round of testing persisted in the more recent round of testing, across the denominator exclusion conditions described above. Based on the quantitative testing results and input from experts on the measure´s face validity and usability, the current measure specification includes the following denominator exclusion - patients with Tourette´s, Bipolar, Huntington´s, or schizophrenia and/or patients taking antipsychotics prior to admission.

Age. Testing results indicated that antipsychotic use increases, across all denominator exclusion conditions, as age increases. Among patients ages 65 to 74 years, across denominator exclusion conditions, the performance rate ranged from 21.1 to 28.8. Among patients 85 and older, the performance rate ranged from 41.0 to 50.5. This is an important finding as there has been significant concern about the inappropriate use of antipsychotics among older individuals. These findings support the notion that older patients are more likely to receive antipsychotics than younger patients, despite the AGS Beer's Criteria which cautions against their use in older adults. This lends support to the importance of this measure. Chi-squared analyses were done to exam subgroup differences. Across the four denominator exclusion conditions, subgroup differences are statistically significant (p<.001) for age.

Sex, race, and ethnicity. Males had higher rates of antipsychotic ordering than females across all denominator exclusion conditions. Across conditions, performance rates among males ranged from 34.2 to 42.4. For females, the range was 23.4 to 32.6. With regard to race, Blacks were more likely than Whites to be ordered an antipsychotics, across all denominator exclusion conditions. Hispanic and non-Hispanic patients had similar performance rates across denominator exclusion conditions. Chi-squared analyses were done to exam subgroup differences. Across the denominator exclusion conditions, subgroup differences are statistically significant (p<.001) for sex and race. There were no significant difference in performance scores by ethnicity in any of the denominator exclusion conditions.

Payer. Patients with Medicare, 92 percent of encounters) and Medicaid (1.3 percent of encounters) had the highest rate of antipsychotic ordering across denominator exclusion conditions. As expected, those with private insurance had the lowest performance rates. This is expected as most are younger and covered by insurance through their employers. Chi-squared analyses were done to exam subgroup differences. Across the four denominator exclusion conditions, subgroup differences are statistically significant (p<.001) for payer.

Data collected during measure testing on all adult patients (18 and older) found that patients with Medicare and Medicaid coverage had the highest rate of antipsychotic ordering (21.0 percent and 20.1 percent, respectively). Patients with private insurance had the lowest rates at 11.3 percent. Measure testing found statistically significant differences in antipsychotic orders for patients 65 and older compared to patients age

18–64 (21.6 versus 14.8, respectively) and significant differences in males compared to females (19.8 and 16.1 percent, respectively). There was little difference in the rate of antipsychotic ordering by race or ethnicity. Across racial groups (black, white, and other), the rate of antipsychotic ordering ranged from 17.6 to 18.4. Hispanic and non-Hispanic patients had similar performance rates, 16.6 and 18.1 respectively. Although these differences are small and likely not clinically significant, they are statistically significant (p<.001). This is likely due to the large sample size. (Barrett 2017).

Barrett, K,. F. Xing, K. Sobel, and B. Rehm. "Hospital Inpatient and Outpatient Process and Structural Measure Development and Maintenance Project: Beta Testing Report on the Use of Antipsychotics in Adults in the Inpatient Hospital Setting Electronic Clinical Quality Measure." Washington, DC: Mathematica Policy Research, July 2017.

Barrett, K,. F. Xing, K. Sobel, and B. Rehm. "Hospital Inpatient and Outpatient Process and Structural Measure Development and Maintenance Project: Beta Testing Report Addendum on the Use of Antipsychotics in Adults in the Inpatient Hospital Setting Electronic Clinical Quality Measure." Washington, DC: Mathematica Policy Research, August 2019.

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

The research on disparities in the use of antipsychotics is limited. According to one researcher, factors such as insurance status and race have been associated with the use of antipsychotics in hospitalizations. Patients with Medicare, Medicaid, or self-pay primary insurance are more likely to receive antipsychotics than patients with commercial primary insurance (Herzig, 2016a). Herzig also observed that non-white individuals are less likely to receive antipsychotics than white individuals. Further scientific investigation is required to understand the reasons for these disparities (Herzig 2016b).

Herzig, S. J., M. B. Rothberg, et al. (2016a). "Antipsychotic medication utilization in nonpsychiatric hospitalizations." J Hosp Med.

Herzig, S. J., M. B. Rothberg, et al. (2016b). "Antipsychotic Use in Hospitalized Adults: Rates, Indications, and Predictors." J Am Geriatr Soc 64(2): 299-305.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

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

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

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

Not applicable

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

This is an eMeasure **Attachment:** CMS498_v5_7_Artifacts-637025216008122695.zip,BonnieTestPatientExport_CMS498v0-637025216008122695.xlsx,1a._AP_Logic_Flow-637025216008122695.pdf

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

Attachment **:** AP_value_sets_codes.xlsx

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

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

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

Not an instrument-based measure

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

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

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

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

Inpatient hospitalizations for patients who received an order for an antipsychotic medication during the inpatient encounter.

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

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

The time period for data collection is the measurement year (12-month period).

Numerator: Inpatient hospitalizations for patients who received an order for an antipsychotic medication during the inpatient encounter.

Antipsychotic orders are represented with the QDM datatype and value set of Medication, Order: Antipsychotic Medications (OID:2.16.840.1.113883.3.464.1003.196.12.1255).

Numerator exclusions: Inpatient hospitalizations for patients with documented indication that they are threatening harm to self or others.

Threat to self or others is represented with the QDM datatype and value set of Symptom: Threat to themselves or others (OID:2.16.840.1.113883.3.464.1003.195.12.1020).

To access the value sets for the measure, please visit the Value Set Authority Center, sponsored by the National Library of Medicine, at https://vsac.nlm.nih.gov/. A list of value sets for the measure is attached in the Excel workbook provided for question S.2b.

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

Non-psychiatric inpatient hospitalizations for patients who are 65 and older.

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

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

The time period for data collection is the measurement year (12-month period).

Denominator: Non-psychiatric inpatient hospitalizations for patients who are 65 and older.

Inpatient hospitalizations are represented with the QDM datatype and value set of Encounter, Performed: Encounter Inpatient (OID:2.16.840.1.113883.3.666.5.3001).

To access the value sets for the measure, please visit the Value Set Authority Center, sponsored by the National Library of Medicine, at https://vsac.nlm.nih.gov/. A list of value sets for the measure is attached in the Excel workbook provided for question S.2b.

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

Inpatient hospitalizations for patients with a diagnosis of schizophrenia, Tourette´s syndrome, bipolar disorder, Huntington´s disease during the encounter.

Inpatient hospitalizations for patients who were taking antipsychotics prior to admission.

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

The following data elements are used to define the measure exclusions:

Denominator Exclusions: Inpatient hospitalizations for patients with a diagnosis of schizophrenia, Tourette´s syndrome, bipolar disorder, Huntington´s disease during the encounter. These exclusions are represented with the QDM datatype of Diagnosis.

Schizophrenia or Psychotic Disorder (OID: 2.16.840.1.113883.3.464.1003.105.12.1104)

Tourette´s Syndrome (OID: 2.16.840.1.113883.3.464.1003.105.12.1030)

Bipolar Disorder (OID: 2.16.840.1.113883.3.67.1.101.1.128)

Huntington´s Disease (OID: 2.16.840.1.113883.3.464.1003.105.12.1032)

Denominator Exclusions: Inpatient hospitalizations for patients who were taking antipsychotics prior to admission.

Antipyschotic Medications (OID: 2.16.840.1.113883.3.464.1003.196.12.1255)

This exclusion is represented with the QDM datatype of Medication, Active:

Antipsychotic Medications (OID: 2.16.840.1.113883.3.464.1003.196.12.1255)

To access the value sets for the measure, please visit the Value Set Authority Center, sponsored by the National Library of Medicine, at https://vsac.nlm.nih.gov/. A list of value sets for the measure is attached in the Excel workbook provided for question S.2b.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

Results include a total score and the following strata:

Stratum 1 - Patients who were admitted or transferred to the ICU during the inpatient encounter

Stratum 2 - Patients who were not admitted or transferred to the ICU during the inpatient encounter

These strata are identified using the QDM datatype of Encounter, Performed.

ICU Admission or Transfer (OID: 2.16.840.1.113883.17.4077.3.2040)

To access the value sets for the measure, please visit the Value Set Authority Center, sponsored by the National Library of Medicine, at https://vsac.nlm.nih.gov/. A list of value sets for the measure is attached in the Excel workbook provided for question S.2b.

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

Stratification by risk category/subgroup

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

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

See ´1a._AP_Logic_Flow.pdf´ submitted as an attachment under S.2a above.

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

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

Not applicable

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

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

Not applicable

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

If other, please describe in S.18.

Electronic Health Records

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

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

Hospitals collect EHR data using certified electronic health record technology (CEHRT). The human readable format and XML are contained in the eCQM specifications attached in question S.2a. No additional tools are used for data collection for eMeasures.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Facility

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Inpatient/Hospital

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

Not applicable

2. Validity – See attached Measure Testing Submission Form

HMDM_Testing_form_2019_07_26_AP_Resubmit-637025216010935328.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

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

Measure Number (*if previously endorsed*)**:** Not applicable
Measure Title: Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting
Date of Submission: 11/1/2017 (original); XX/X/XXXX (resubmission)

Type of Measure:

☐ **Outcome (*including PRO-PM*)**	☐ **Composite – *STOP – use composite testing form***

☐ Intermediate Clinical Outcome	☐ Cost/resource
☒ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

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

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

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

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

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

Not applicable. We did not use an existing data set to test this measure; instead, to test this measure, we partnered with three test sites to extract data from their EHR systems (described in question 1.5). In alignment with the measure's general intent of assessing the use of antipsychotics, we asked hospital staff to submit patient-level data for all patients that qualify for the initial patient population over a one- to two-year period, which includes inpatient admissions for patients 18 years and older (as of the date of the encounter), excluding those with a principal diagnosis of Huntington's, Tourette's, bipolar, or schizophrenia, and where these medications are FDA approved for use. Since the measure is specified for older adults ages 65 and above, all analyses provided in this test report are limited to that age cohort.

2019 update for resubmission: In November 2017, the NQF Behavioral Health Committee reviewed this submission for endorsement consideration for the Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting measure ('antipsychotic measure'). During the Behavioral Health Committee meeting, members requested that the measure development team consider two additional exclusions – antipsychotic use prior to admission and antipsychotic use for treatment resistant depression. To respond to concerns raised by NQF's Behavioral Health Committee, we requested that Test Site 1, consisting of 9 hospitals, create a new extract with the patient-level data described above with two additional data elements for the exclusions. One exclusion flagged antipsychotics prescribed prior to admission and a second exclusion flagged patients with a primary or secondary diagnosis of treatment-resistant depression. In addition, to allow for an assessment of data element validity, Test Site 1 also manually abstracted data for 200 randomly selected encounters to

confirm validity of the new data elements. The results of testing are contained within this document and are highlighted in blue for ease of review.

1.3. What are the dates of the data used in testing? 10/1/2013-9/30/2015 (Test Site 1 and 2); 10/1/2014 – 9/30/2015 (Test Site 3)

2019 update for resubmission: We partnered with Test Site 1 to access patient-level data for the time period January 1, 2018 thru December 31, 2018.

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

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

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

Our test data includes data from 11 hospitals across three test sites (two health systems and one critical access hospital). When selecting sites, we ensured representation of at least two different EHR systems across sites, as required by NQF for eCQM testing. We also purposely sought sites whose EHR systems captured the data elements required for the measure calculation and had the ability to create an electronic data extract. By selecting test sites that could provide data for multiple hospitals, we were able to achieve a mix of urban and rural settings and care settings with large and small bed counts. All test sites were non-profit. Test Sites 1 and 2 are teaching hospitals. Test Site 3 is a small, rural safety net hospital. Table 1 lists characteristics of the hospitals participating in field testing.

2019 update for resubmission: We partnered with Test Site 1 to access patient-level data for the same 9 hospitals from which we received patient-level data during initial testing. The characteristics of Test Site 1 in Table 1 remained the same for the 2019 update for resubmission.

Table 1. Field testing hospital characteristics

	Hospital	State	Geography	# of beds	EHR product	Inception of current EHR system
Test site 1	All	TX	Urban	3,320	Cerner	2006
	1	TX	Urban	260	Cerner	2006
	2	TX	Urban	877	Cerner	2006
	3	TX	Urban	142	Cerner	2006
	4	TX	Urban	444	Cerner	2006
	5	TX	Urban	255	Cerner	2006
	6	TX	Urban	274	Cerner	2006
	7	TX	Urban	149	Cerner	2006
	8	TX	Urban	568	Cerner	2006
	9	TX	Urban	351	Cerner	2006
Test site 2	10	NC	Urban	874	Cerner	2006
Test site 3	11	PA	Rural	25	Meditech	2010

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)
Across the three test sites, we received data for 58,507 patient encounters. Across sites, the average age of patients was 76.5 years with a range across sites from a low of 74.6 years to a high of 78.1 years. Distribution by sex was fairly even for Test Sites 1 and 2; at Test Site 3, approximately two-thirds of the patients were female. Across sites, the majority of patients were White and non-Hispanic. At Test Sites 1 and 2, over 90 percent of patients had Medicare. Approximately 60 percent of patients at Test Site 3 had Medicare and 36 percent had private insurance. See Table 2 for a breakdown of these demographic characteristics by test site.

Table 2. Demographic characteristics of the field-testing sample

Characteristics	Test site 1		Test site 2		Test site 3		Across sites (pooled data)	
	N	%	N	%	N	%	N	%
Number of patients	45,097		12,954		456		58,507	
Average age	77.0		74.6		78.1		76.5	
Sex								
Male	18,948	42.0	6,277	48.5	172	37.7	25,397	43.4
Female	26,145	58.0	6,677	51.5	284	62.3	33,106	56.6
Race								
White	28,381	65.7	9,810	76.6	455	99.8	38,446	68.3
Black	6,407	14.8	2,543	19.9	1	0.2	8,951	15.9
Other	8,439	19.5	457	3.6	0	0.0	8,896	15.8
Ethnicity								
Hispanic	4,889	10.8	175	1.4	0	0	5,064	8.7
Non-Hispanic	37,215	82.5	12.404	95.8	436	95.6	50,055	85.6
Other or unknown	2,993	6.7	375	2.8	20	4.4	3,388	5.8
(Primary) Payer								
Medicare	41,415	91.8	11,924	92.1	279	61.2	53,618	91.6
Medicaid	585	1.3	83	.64	9	2.0	677	1.2
Private insurance	2,513	5.6	764	5.9	166	36.4	3,443	5.9
Self-pay or uninsured	283	0.6	63	0.5	0	0.0	346	0.6
Others	301	0.7	120	0.9	2	0.4	423	0.7

SOURCE: Test Sites 1 and 2 data from October 1, 2013 to September 30, 2015. Test Site 3 from October 1, 2014 to September 30, 2015.

2019 update for resubmission: We received data for 16,621 patient encounters. Patient encounter characteristics at Test Site 1 were similar to those seen in the initial round of testing, as seen in Table 2a. Distribution by race differed between initial testing and testing for the 2019 resubmission, with a higher proportion of patients classified as 'other' in the resubmission (38 percent vs. 20 percent) and a smaller proportion of patients classified as 'white' (49 percent vs. 66 percent).

Table 2a. Demographic characteristics of the field-testing sample – initial testing and 2019 resubmission

Characteristics	Test site 1 Original Testing		Test site 1 2019 Resubmission	
	N	%	N	%
Number of patient encounters	45,097	100.0	16,621	100.0
Average age	77.0	---	77.3	---
Sex				
Male	18,948	42.0	7,019	42.2
Female	26,145	58.0	9,598	57.7
Unknown	-----	----	4	.02
Race				
White	28,381	65.7	8,113	48.8
Black	6,407	14.8	2,221	13.4
Other	8,439	19.5	6,287	37.8
Ethnicity				
Hispanic	4,889	10.8	1,585	9.5
Non-Hispanic	37,215	82.5	14,285	85.9
Other or unknown	2,993	6.7	751	4.5
(Primary) Payer				
Medicare	41,415	91.8	15,310	92.1
Medicaid	585	1.3	285	1.7
Private insurance	2,513	5.6	775	4.7
Self-pay or uninsured	283	0.6	120	.7
Others	301	0.7	131	.8

SOURCE: Test Sites 1 data from January 1, 2018 to December 31, 2018.

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

- Reliability: We used electronically extracted EHR data from 11 hospitals to examine the reliability of the measure performance rate. Data used were for the time period described in Question 1.3.
- Data element validity: We randomly selected a sample of encounters from each test site's electronic EHR extract and manually abstracted data for those encounters in order to assess the chance-adjusted agreement between the two sources. Manual abstraction was done by trained medical record abstractors. A total of 158 encounters were abstracted across test sites.
- Face validity. We solicited feedback on face validity via interviews and a brief web survey from clinicians, information technology professionals, subject matter experts, and members of the expert workgroup (n=8 respondents).
- Exclusions: We used electronically extracted EHR data from 11 hospitals to examine the impact of the numerator and denominator exclusions on the measure's performance rate. Data used were for the time period described in Question 1.3.
- Risk adjustment: Not applicable; this measure is not risk adjusted.

- Meaningful difference in performance: We used electronically extracted EHR data from 11 hospitals to identify difference in performance by test sites and by demographic characteristics such as age, race, gender, sex, and payer source. Data used were for the time period described in Question 1.3.
- Missing data/bias: We used electronically extracted EHR data from 11 hospitals to examine the extent to which age and admission and discharge dates were missing in the electronically extracted data from test sites' EHR. Data used were for the time period described in Question 1.3.

2019 update for resubmission:

- Reliability: We used electronically extracted EHR data from 9 hospitals at Test Site 1 to examine the reliability of the measure performance rate. Data used were for hospital discharges between January 1, 2018 and December 31, 2018).
- Data element validity: We randomly selected a sample of encounters from Test Site's 1 electronic EHR extract and manually abstracted data for those encounters in order to assess the chance-adjusted agreement between the two sources. Manual abstraction was done by trained medical record abstractors. A total of 200 encounters were abstracted.
- Face validity: We conducted in-depth interviews with representatives from two hospital systems and two members of the Antipsychotics Measure Development EWG to gather input on the face validity of the two additional exclusion criteria – antipsychotics prior to admission and antipsychotics with a treatment-resistant depression diagnosis.
- Exclusions: We used electronically extracted EHR data from 9 hospitals at Test Site 1 to examine the impact of two additional denominator exclusions on the measure's performance rate (antipsychotics prior to admission and antipsychotics with a treatment-resistant depression diagnosis). Data used were for hospital discharges between January 1, 2018 and December 31, 2018).
- Meaningful difference in performance: We used electronically extracted EHR data from 9 hospitals at Test Site 1 to identify difference in performance by hospital. Data used were for hospital discharges between January 1, 2018 and December 31, 2018).
- Missing data: Data element validity (comparison of chart with EHR extracted data) provides a comprehensive assessment of missingness. If data are missing from the EHR extracted data but not the chart abstract data element validity will be low.

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

As described in section 1.6, we collected information on the following variables using data extracted from hospital EHR systems: age, sex, race, ethnicity, and payer. This measure is based on a process that should be carried out for all patients (except those excluded), so no adjustment for patient mix is necessary. We did collect information about these five variables and assessed disparities in performance rate for each group. Those results are described in section 2b5.

2019 update for resubmission: Same as above.

2a2. RELIABILITY TESTING

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

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

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

We employed the split-half correlation to assess the reliability of the performance measure scores. The split-half correlation is a way to implement the test-retest reliability method. It estimates the measure reliability directly from the data and is less constrained by a small number of test sites than other model-based methods that require more data to justify model assumptions (for example, signal-to-noise using Beta-binomial model). The split-half correlation characterizes the correlation of estimated measure results between two non-overlapping data sets. To estimate the reliability, we randomly divided the hospital-level EHR data into two equal samples. We then calculated the measure performance in both samples for each hospital and calculated the weighted correlation between the estimates of the performance rate (the hospital's weight is based upon its number of denominator cases to account for the sample size effect in each hospital). The higher the correlation, the higher the statistical reliability of the measure. Stated another way, the higher the correlation, the greater the amount of variation that can be explained through systematic differences across the test sites as opposed to random error (for example, sampling variation within measured entities). To produce more stable estimates, we repeated this resampling approach more than 2,500 times. We used 0.4 as our benchmark level for an acceptable estimate of measure reliability because it aligns with guidance in the literature; Evans (1996) suggests that for the absolute value of Pearson's correlation *r*, a range of 0.40–0.59 indicates "moderate" reliability.

[Reference: Evans, J. D. (1996.) *Straightforward Statistics for the Behavioral Sciences*. Brooks/Cole Publishing, Pacific Grove.]

2019 update for resubmission: Same as above.

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

Reliability tests were conducted, as described in section 2a2.2, to generate a reliability score for the measure. Because we are looking at measure-level reliability, the measure has one reliability score:

Table 3. Reliability testing results

Measure name	Reliability score	95% Confidence Interval
Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting, 65 years of age and older	0.981	0.957, 0.995

2019 update for resubmission: Reliability tests were conducted, as described in section 2a2.2, to generate a reliability score for the measure as originally specified and for the measure under three additional scenarios – excluding encounters with antipsychotics prior to admission, excluding encounters with antipsychotics with a treatment-resistant depression diagnosis, excluding encounters with antipsychotics prior to admission or antipsychotics with a treatment-resistant depression diagnosis. Table 3a provides the results.

Table 3a. Reliability testing results – 2019 resubmission

Measure name	Reliability score	95% Confidence Interval
Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting, 65 years of age and older (original specification)	.95	.89, .99
Original specification plus exclusion of patients taking antipsychotics prior to admission	.95	.89, .99
Original specification plus exclusions of patients with a treatment-resistant depression diagnosis and antipsychotic use	.95	.88, .99
Original specification plus exclusion of patient taking antipsychotics prior to admission and/or patients with a treatment-resistant depression diagnosis and antipsychotic use	.95	.88, .99

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

We assessed performance rate reliability across test sites using a split-half correlation. The reliability coefficient across 11 hospitals for the antipsychotic measure was .98 (with a 95 percent confidence interval, (0.96, 0.99) for all encounters, 65 years of age and older. This indicates that the hospital-level performance rate has excellent reliability, and is relatively free from measurement error. Reliability coefficients of .9 or above reflect excellent precision between performance rates derived from the two samples (a reliability coefficient of 1.0 reflects perfect precision).

[Reference: Adams, John L. "The Reliability of Provider Profiling: A tutorial." Santa Monica, CA: RAND Corporation, 2009.]

2019 update for resubmission: We assessed performance rate reliability across the 9 hospitals at Test Site 1 using a split-half correlation. The reliability coefficient across the 9 hospitals for the antipsychotic measure, as originally specified, was .95 (with a 95 percent confidence interval, (0.89, 0.99) for all encounters, 65 years of age and older. The reliability coefficient for the three additional conditions, antipsychotics prior to admission, antipsychotics for treatment resistant, and antipsychotics prior to admission and/or antipsychotics with a

treatment-resistant depression diagnosis were .95 (.89,.99), .95 (.88,.99), and .95 (.88,.99) respectively. This indicates that the hospital-level performance rate has excellent reliability across the different measure conditions, and is relatively free from measurement error. Reliability coefficients of .9 or above reflect excellent precision between performance rates derived from the two samples (a reliability coefficient of 1.0 reflects perfect precision).

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☐ **Empirical validity testing**

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

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

Data element (Criterion) Validity

Data element validity testing evaluated whether the measure specification correctly identifies all the data elements required to calculate the measure score. This method quantifies the percent agreement, Kappa statistic, sensitivity, specificity, negative predictive value and positive predictive value between the electronically extracted EHR data and the manually abstracted data (which use the entire record, including free text notes fields). Each of these statistics illustrates the closeness between data element results from the two sources. In general, the higher the value, the more consistency between the data from the two sources.

Data element validity was tested by selecting a random set of patient encounters from the full electronic EHR extract and comparing data for these encounters to those that were manually abstracted, by trained abstractors, for the same encounters. The manually abstracted EHR data were considered the 'gold standard' against which we assessed the validity of the EHR-extracted data.

2019 update for resubmission: The same procedures were used in testing for the resubmission. Analyses were limited to the two new data elements – antipsychotics prior to admission and antipsychotics with a treatment-resistant depression diagnosis.

Face Validity

Formal Assessment of Face Validity (EWG and staff at test sites)

We evaluated the face validity of the measure specification and the measure score by surveying eight experts via the web: two clinicians from Test Sites 1 and 3, four Expert Work Group (EWG) members (three physicians, one academic), and two quality improvement / informatics staff from Test Sites 1 and 2. The survey asked respondents about the appropriateness of the measure components (denominator, denominator exclusions, numerator, and numerator exclusions) given the intent of this measure. In addition, we asked respondents if hospitals that 1) document "threat of harm" for patients that are prescribed antipsychotics, and 2) document denominator exclusions, should score well on the measure. For each item, respondents indicated the extent to which they agreed (1 = Strongly agree; 2 = Agree, 3 = Disagree; 4 = Strongly disagree).

The EWG, which included physicians, academicians, and subject matter experts, helped ensure that the measure specification and measure score have a high degree of face validity. EWG members are listed in Table 4. We also evaluated the face validity of the measure specification and the measure score by soliciting input from key stakeholders during public comment, the Patient and Family Advisory Board (PFAB) and the Technical Expert Panel.

Table 4. EWG Members

Name	Organization
Byron Bair, MD, MBA	Salt Lake City VA
Soo Borson, MD	University of Washington
Josh Chodosh, MD, MSHS	NYU School of Medicine
Elizabeth Galik, RN, PhD, CRNP	University of Maryland School of Nursing
Susan Merel, MD	University of Washington Department of Medicine
Paul Rosenberg, MD	Johns Hopkins
Lynn Shell, PhD, APN, CARN-AP	Rutgers
Teepa Snow, MS, OTR/L, FAOTA	Positive Approach, LLC
Heidi Wald, MD, MS, MS	University of Colorado

2019 update for resubmission: We conducted in-depth interviews with representatives from two hospital systems and two members of the Antipsychotics Measure Development EWG to gather input on the face validity of the two additional exclusion criteria (antipsychotics prior to admission and use of antipsychotics with a treatment-resistant depression diagnosis

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

Data Element Validity

There were high levels of agreement between the electronically extracted and manually abstracted EHR data for the denominator, denominator exclusions, numerator, and numerator exclusions across the three test sites. Table 5 describes the level of agreement between the two data sources for each component of the measure specification. The chart-abstracted data represent the gold standard for data element validity testing.

Table 5. Agreement statistics for random sample data between EHR extraction and manual chart abstraction (n=158)

Measure Component	Agreement (%)	Kappa	Sensitivity	Specificity
Denominator	98.1	0*	1	0
Initial Population	98.1	0*	1	0
Denominator exclusion				
Schizophrenia	99.4	0.66	0.88	0.99
Huntington's	100.0	NaN	NaN	1
Bipolar	98.8	0.49	0.5	0.99
Tourette's	100.0	NaN	NaN	1
Numerator (antipsychotic order during encounter)	100.0	1.0	NaN	1
Numerator exclusion	98.1	0.39	0.33	0.99

Source: Data from 10/1/2013 to 9/30/2015 for Test Sites 1 and 2, and 10/1/2014 to 9/30/2015 for Test Site 3.
Notes: NaN: Not calculable because the denominator in the equation is equal to zero.

*All 158 cases were contained within the denominator from the EHR. Chart abstractors flagged 3 of the 158 cases as not meeting denominator criteria. The Kappa statistic treats the 155 yes-yes agreement largely as "chance agreement" and penalizes this condition when applying the chance correction.

We measured overall agreement, defined as the number of patients for which both sources agree on the presence or absence of a condition among all patients tested. We also used Cohen's Kappa statistic to reflect chance-adjusted agreement. The Kappa score can range from -1.00 to 1.00. Although higher Kappa scores tend to indicate higher agreement between two data sources, a low Kappa score may not represent low agreement when the data are imbalanced.

The overall sample of 158 encounters showed 98 percent agreement or higher for all data elements and data element combinations assessed. In addition, agreement was perfect for two of the exclusionary data elements (Tourette's and Huntington's) and the numerator data element (antipsychotic prescription) and almost perfect for the remaining data elements. Kappa values ranged from a low of .39 for the numerator exclusion ("threat of harm") to a high of 1.0 for the numerator (medication orders). The numerator exclusion sensitivity is reflective of the inconsistent documentation of the numerator exclusion ("threat of harm") in the EHR.

[Reference: Viera, Anthony J., and Joanne M. Garrett. "Understanding Interobserver Agreement: The Kappa Statistic." Family Medicine, vol. 37, no.5, 2005, pp. 360–363.]

2019 update for resubmission: Across the 200 manually abstracted encounters, 21 had an antipsychotic ordered and had a diagnosis of treatment-resistant depression, 179 did not. Agreement with the electronically extracted data was 98.5 percent, meaning the electronic data matched the gold-standard abstracted data in 197 of the 200 encounters. In terms of antipsychotics prior to admission, agreement was lower. There were 36 encounters with an antipsychotic prior to admission in the manually abstracted data and 163 without (in the abstraction file, for one record, the antipsychotic field was blank for a total n of 199 valid responses). In the electronic extract, 180 encounters matched for an agreement rate of 90.5 percent.

Table 5a describes the level of agreement between the two data sources for the two additional data elements, as well as measure-level validity when patients with antipsychotics prior to admission are excluded, measure-level validity when patients with antipsychotics with a treatment-resistant depression diagnosis are excluded, and measure-level reliability with either condition met. The chart-abstracted data represent the gold standard for data element validity testing. Kappa, chance-adjusted agreement, improved with additional denominator exclusions. Kappa increased from .31 to .38 when excluding encounters with antipsychotics prior to admission and from .31 to .35 when excluding encounters with antipsychotics and a treatment-resistant depression diagnosis.

Table 5a. Agreement statistics for random sample data between EHR extraction and manual chart abstraction (n=200) – 2019 resubmission

Component	Agreement (%)	Kappa	Sensitivity	Specificity
Data elements				
Antipsychotics with a treatment-resistant depression diagnosis	98.5	.91	.86	1.0
Antipsychotics prior to admission	90.5	.66	.67	.96
Measure				
Use of Antipsychotics in Older Adults in the Inpatient Hospital Setting, 65 years of age and older (original specification)	69.5	.31	.96	.66
Original specification *plus exclusion of patients taking antipsychotics prior to admission*	78.5	.38	.91	.77
Original specification *plus exclusions of patients with treatment-resistant depression and antipsychotic use*	75.5	.35	.95	.73
Original specification *plus exclusion of patient taking antipsychotics prior to admission and/or patients with a treatment-resistant depression diagnosis and antipsychotic use*	81.5	.38	.89	.81

Source: Data from 1/1/2018 to 12/31/2018 for Test Sites 1.

Face Validity

Results from the web-based survey of members of the measure's EWG, and test site representatives indicate that the measure had strong face validity. All respondents (n=8) strongly agreed or agreed that the measure components (denominator, denominator exclusions, numerator, and numerator exclusions) were appropriate to the intent of this measure (stated at the beginning of the executive summary). Further, six out of eight

respondents agreed that hospitals should score well on the measure if they 1) document "threat of harm" for patients that are prescribed antipsychotics, and 2) document denominator exclusions.

Twenty-two comments were received during the 30-day public comment period which ran from April 15, 2016 through May 15, 2016. Commenters included hospitals and health systems (6), professional associations (7), EHR vendors (2), academic institutions (3), and individuals (2). Responses reinforced the measure's main goal of calling attention to off-label antipsychotic prescribing practices, thereby reducing inappropriate use of antipsychotics. Many commenters acknowledged the importance of developing a hospital measure that addresses use of antipsychotics in the inpatient setting. Some, however, highlighted the potential unintended consequences of the measure's implementation in critical care settings. Some of the commenters expressed concern over the overall intent of the measure, suggesting that the measure might unintentionally encourage the use of potentially harmful and less effective alternatives such as benzodiazepines. Also cited was the potential for increased use of physical restraints as an alternative to antipsychotics. These commenters also questioned the ability of the measure to address either appropriate use of antipsychotics or quality care gaps.

The PFAB believed that the antipsychotic measure is important and that it could be used to help decrease the use of antipsychotics during hospitalization and, possibly, long-term. They believed the measure may facilitate proactive provider education and improved hospital policies on managing patient agitation. In addition, the measure may result in greater levels of engagement with the patient as well as his/her family.

The TEP was in agreement about the importance of the measure. There was concern about the measure being focused on medications ordered rather than medications administered. The intent of the measure is to change prescribing behaviors. In the future, CMS may consider adding a second numerator for antipsychotics administered.

2019 update for resubmission: We conducted in-depth interviews with representatives from two hospital systems and two members of the Antipsychotics Measure Development EWG to gather input on the face validity of the two additional exclusion criteria (antipsychotics prior to admission and use of antipsychotics with a treatment-resistant depression diagnosis). All of those interviewed indicated that excluding patients using antipsychotics prior to admission is appropriate because 1) clinically, stopping antipsychotics that are in use prior to admission could have deleterious consequences for patients and 2) hospitals should not be held accountable for antipsychotic use when it is being managed by an out-patient provider. In terms of the antipsychotics with a treatment-resistant depression diagnosis, those interviewed reported that antipsychotics are not typically used for the management of treatment-resistant depression and, as a result, the number of patients impacted by this exclusion would be minimal. For this reason, the consensus among those interviewed was that adding an exclusion for antipsychotics with a treatment-resistant depression diagnosis was not warranted.

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

The Kappa values calculated through data element validity testing suggest that data in the EHR accurately reflect patient care. In addition, face validity appears to be high as well. Six out of eight respondents reported that hospitals would score well on the measure if they consistently documented "threat of harm" and denominator exclusions. One person who disagreed commented that the denominator exclusions should be broader, noting that there were other diagnoses for which patients were on chronic antipsychotics. The other respondent who disagreed did not provide qualitative feedback.

2019 resubmission:

Data element validity, measured using kappa agreement between EHR extracted data and chart abstracted data for a set of patients, improved with the addition of two exclusions recommended by the review panel from 0.31 (original specification) to 0.38 (adding antipsychotics prior to admission) and 0.35 (adding antipsychotics with a treatment-resistant depression diagnosis).

In terms of face validity, based on input from experts and clinicians, antipsychotics prior to admission is an appropriate denominator exclusion. However, those interviewed did not agree with having antipsychotics for the treatment of depression as a denominator exclusion because antipsychotics are not commonly used to treat that condition and the number of patients impacted would be very small.

2b2. EXCLUSIONS ANALYSIS

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

The following five exclusions apply to the measure. Excluded from the denominator are encounters with a documented diagnosis of schizophrenia, Huntington's, bipolar disorder, or Tourette's. These are conditions for which antipsychotics are approved for use. Excluded from the numerator are encounters with a documented "threat of harm to self or others." This exclusion is supported by clinical guidelines.

2019 resubmission: Two additional denominator exclusions were considered, antipsychotics prior to admission and antipsychotics with a treatment-resistant depression diagnosis.

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

To examine the effect of the exclusions, the number affected by exclusions was first examined and the measure rates with and without each exclusion were calculated and compared.

2019 resubmission: Same as above.

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

Table 6 shows the count of cases within each exclusion category across sites and within sites. It should be noted that within an encounter, a patient may have more than one exclusion.

Table 6. Number and proportion of exclusions

	Test site 1		Test site 2		Test site 3		Across sites (pooled data)	
	N	%	N	%	N	%	N	%
Number of encounters	45,097	100.0	12,954	100.0	456	100.0	58,507	100.0
Number of encounters in denominator exclusion	1,316	2.9	310	2.4	1	0.2	1,627	2.8
Number of encounters in numerator exclusion	104	0.2	48	0.4	0	0.0	152	0.3

Table 7 shows performance rates by test site for the measure as it is currently specified with exclusions (Column A), the measure with no numerator exclusion (Column B), the measure with no denominator exclusions included in the calculation (Column C), and the measure including one of the four denominator exclusions (schizophrenia, Huntington's Disease, bipolar disorder, Tourette's Syndrome) (Columns D, E, F, G, respectively).

Table 7. Comparison of performance rate based on exclusion criteria

	COLUMN A Performance rate. Measure as specified	COLUMN B Performance rate without numerator exclusion	COLUMN C Performance rate without denominator exclusions	COLUMN D Performance rate including schizophrenia; excluding Huntington's, bipolar, and Tourette's	COLUMN E Performance rate including Huntington's; excluding schizophrenia, bipolar, and Tourette's	COLUMN F Performance rate including bipolar; excluding schizophrenia, Huntington's, and Tourette's	COLUMN G Performance rate including Tourette's; excluding schizophrenia, Huntington's, and bipolar
Total	21.6	21.7	22.6	22.0	22.5	21.9	22.6
Test site 1	21.5	22.7	22.5	22.0	22.5	21.8	22.5
Test site 2	22.5	22.7	23.4	22.9	23.4	23.0	23.4
Test site 3	6.6	6.6	6.8	6.8	6.8	6.6	6.8

2019 resubmission: Table 6a shows the count of cases within each exclusion category at Test Site 1 as well as the performance rate. It should be noted that within an encounter, a patient may have more than one exclusion. Approximately 10 percent of encounters are removed from the denominator when applying the antipsychotic with a treatment-resistant depression diagnosis and 16 percent are removed when applying the antipsychotics prior to admission exclusion.

Table 6a. Number and proportion of denominator exclusions

	Encounters after exclusions	%	Performance Rate
Total number of encounters (IPP)	16621	100.0	-----
Denominator exclusions			
Total encounters in denominator after excluding encounters where patients have Tourette's, Huntington's, schizophrenia, or bipolar disorder (original denominator exclusions)	15697	94.4	36.8

Total encounters in denominator after original denominator exclusion AND exclusion of encounters where patient was taking antipsychotics prior to admission (regardless of treatment resistant depression status)	14074	84.7	30.3
Total number of encounters after original denominator exclusion AND exclusion of encounters where patient was taking antipsychotic with a treatment-resistant depression diagnosis (regardless of antipsychotic prior to admission status)	14980	90.1	33.9
Total number of encounters after original denominator exclusion AND exclusion of encounters where patient was taking antipsychotic with depression diagnosis and/or antipsychotic prior to admission	13603	81.8	28.0

2019 resubmission: Performance rates varied across the four different denominator exclusion conditions. The performance rate was highest when the original specification was used, 36.8 percent. When excluding encounters with antipsychotics prior to admission, the performance rate dropped to 30.3 percent. When excluding encounters with antipsychotics with a treatment-resistant depression diagnosis, the performance rate dropped slightly to 33.9 percent. When both conditions, antipsychotic prior to admission and antipsychotics with a treatment-resistant depression diagnosis were applied, the performance rate was 28.0 percent across all hospitals at Test Site 1.

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)
Performance rates vary little regardless of the denominator or numerator exclusions. When including all patients 65 years of age and older in the denominator regardless of diagnosis, the performance rate increases one percentage point from 21.6 percent (measure as specified) to 22.6 percent. Similarly, if we remove the numerator exclusion and include all patients who received an order for antipsychotics in the measure calculation regardless of "threat of harm" documentation, the rate increases slightly from 21.6 to 21.7 percent. This minimal difference is not surprising since it has been reported that "threat of harm" documentation is often lacking. Based on testing, the results suggest that numerator and denominator exclusions have little impact on the performance rate. However, for face validity, clinician acceptance of the measure, and consistency with clinical guidelines, it is recommended that the measure exclusions remain as specified.

2019 resubmission: Experts and clinicians supported excluding patient encounters in which antipsychotics were being used prior to admission. Based on the qualitative feedback as well as the noticeable decline in the performance rate when excluding antipsychotics prior to admission, it is recommended that this be included as a denominator exclusion. Experts and clinicians did not agree with an additional exclusion for antipsychotics with a treatment-resistant depression diagnosis. Based on the qualitative feedback and the small change in the performance score when the exclusion is added, we recommend not including it as a denominator exclusion.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

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

☒ **No risk adjustment or stratification**

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

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

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

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

Not applicable.

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

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?
Not applicable.

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

- ☐ **Published literature**
- ☐ **Internal data analysis**
- ☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?
Not applicable.

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

Not applicable.

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

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

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
Not applicable.

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Not applicable.

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Not applicable.

2b3.9. Results of Risk Stratification Analysis:

Not applicable.

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable.

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

Not applicable.

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

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

We analyzed the data to determine if there were statistically significant differences in performance rates by hospital or by age, sex, race, ethnicity, or payer. We also examined differences in performance rates based on intensive care unit (ICU) exposure (encounters with an ICU exposure vs. encounters without an ICU exposure).

To identify statistically significant differences in performance across multiple hospitals, we examined the distribution of performance rates across hospitals. In addition, we calculated the 95 percent confidence interval of the performance rate for each hospital using a z-distribution for proportion. Then we compared each hospital's confidence interval to the overall performance rate, which includes all patients across hospitals. Hospitals with confidence intervals higher than the overall rate indicate room for improvement.

In addition, we conducted chi-square tests to test statistically significant differences in performance between disparity groups, and between care settings.

2019 resubmission: Same as above except that performance rates by ICU exposure were not calculated during the second round of testing at Test Site 1, since the focus was on the two additional data elements: antipsychotics prior to admission and antipsychotics with a treatment-resistant depression diagnosis.

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

Performance by hospital

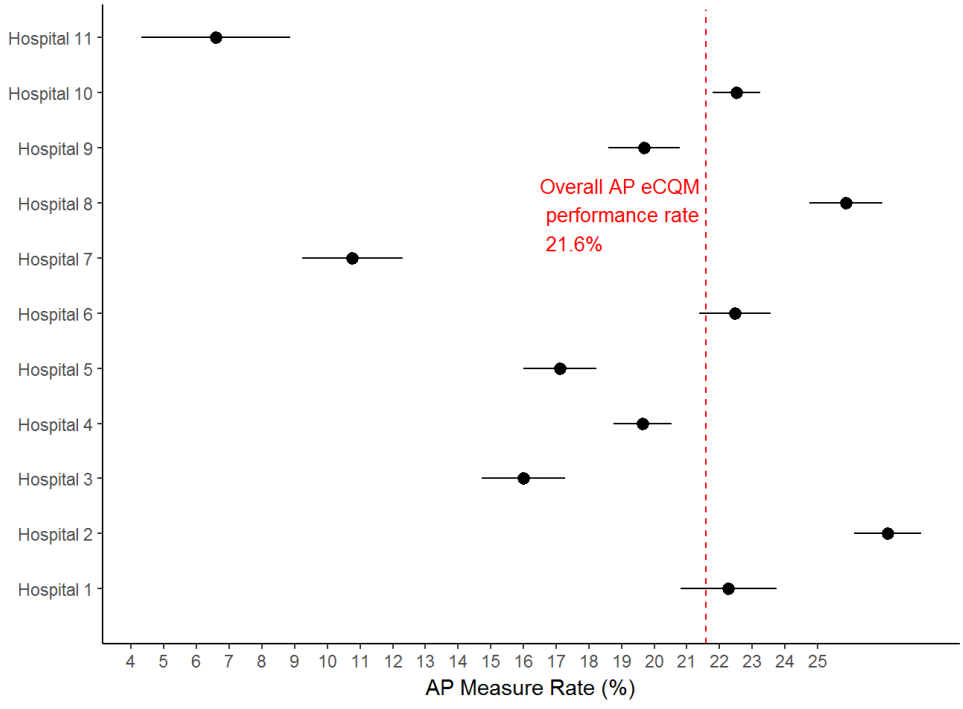
In Table 8, we provide performance rates for each hospital across the three test sites. Performance rates varied from a low of 6.6 percent at Test Site 3 to a high of 25.9 at one of the hospitals in Test Site 1.

Table 8. Antipsychotic electronic clinical quality performance rates (hospital level)

Hospital	Antipsychotic order (%)
Test site 1	
Hospital 1	22.3
Hospital 2	27.1
Hospital 3	16.0
Hospital 4	19.6
Hospital 5	17.1
Hospital 6	22.5
Hospital 7	10.8
Hospital 8	25.9
Hospital 9	19.7
Test site 2	22.5
Test site 3	6.6

The variation in hospital-level measure performance is further illustrated in Figure 1, which shows the distribution of the performance rate and 95 percent confidence interval for each hospital relative to the overall performance rate. The confidence interval is the range in which each hospital's performance rate would likely fall if extractions were repeated multiple times. Although some hospital rates were below the overall performance rate, five out of the 11 hospitals (45.4 percent) have measure rates significantly higher than the overall measure rate (21.6%), indicating room for improvement.

Figure 1. Distribution of performance rates by hospital (orders)



Performance by disparity group

Age. The American Geriatrics Society 2015 Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults cautions against the use of antipsychotics in patients age 65 and older.[1] Testing results indicated that patients ages 65 years and older were more likely to be ordered antipsychotics than patients ages 18 to 64 years , 21.6 percent versus 14.8 percent (not shown in table) (p<.001). Further, when limiting the analysis to those 65 years of age and older, we see a linear relationship with performance rates increasing as age increases (not shown). As seen in Table 9, among patients ages 65 to 74 years, 15.9 percent received an order for antipsychotics compared to 33.0 percent of patients ages 85 years and older (p<.001). This is an important finding as there has been significant concern about the inappropriate use of antipsychotics among older individuals. These findings support the notion that older patients are more likely to receive antipsychotics than younger patients, lending support to the importance of this measure.

Sex, race, and ethnicity. As seen in Table 9, males had higher rates of antipsychotic ordering than females, 24.0 and 19.7 percent, respectively (X^2=154.7, p<.001). With regard to race, Blacks were more likely than Whites to be ordered an antipsychotic, 24.4 versus 20.9 percent, respectively (X^2=54.8, p<.001). There was little difference in the rate of antipsychotic ordering by ethnicity. Hispanic and non-Hispanic patients had similar performance rates, 20.6 and 22.0 percent, respectively (X^2=31.6, p<.001). Although differences based on sex, race, and ethnicity are small and likely not clinically significant, they are statistically significant. This is likely due to the large sample size.

Payer. Patients with Medicare and Medicaid coverage had the highest rates of antipsychotic ordering, 22.0 and 27.9 percent, respectively. Patients with private insurance had the lowest rate at 13.4 percent; this was expected as the measure is focused on older adults (65 years and older). Results were statistically significant (X^2=161.2, p<.001).

[1] The American Geriatrics Society (AGS) Beers Criteria for Potentially Inappropriate Medication (PIM) Use in Older Adults is an explicit list of PIMs best avoided in older adults in general and in those with certain diseases or syndromes, prescribed at reduced dosage or with caution or carefully monitored. It is one of the most frequently consulted sources about the safety of prescribing medications for older adults. The AGS Beers Criteria are used widely in geriatric clinical care, education, and research and in development of quality indicators. Accessed on June 21, 2017 at https://guideline.gov/summaries/summary/49933/american-geriatrics-society-2015-updated-beerscriteria-for-potentially-inappropriate-medication-use-in-older-adults?q=diabetes.

Table 9. Performance rate for antipsychotic ordered measure by patient characteristic

Characteristics	Test site 1		Test site 2		Test site 3		Across sites (pooled data)	
	N	Performance Rate (%)	N	Performance Rate (%)	N	Performance Rate (%)	N	Performance Rate (%)
Number of patients			12,954		456		58,507	21.6
Average age			74.6		78.1		76.5	
Age								
65 to 74			7,414	19.4	191	7.3	27,783	15.9
75 to 84			3,963	24.2	143	3.5	19,307	22.8
85 and older			1,577	32.4	122	9.1	11,417	33.0
Sex								
Male			6,277	25.1	172	5.2	25397	24.0
Female			6,677	20.1	284	7.4	33,106	19.7
Race								
White			9,810	21.7	455	6.6	38,646	20.9
Black			2,543	25.1	1	0.0	8,951	24.4
Other			457	25.1	0		8,896	22.7
Ethnicity								
Hispanic			175	18.8	0		5064	20.6
Non-Hispanic			12.404	22.5	436	6.9	50,055	22.0
Other or unknown			375	24.1	20	0.0	3,388	17.7
(Primary) Payer								
Medicare			11,924	22.5	279	6.8	53,618	22.0
Medicaid			83	25.3	9		677	27.9
Private insurance			764	20.2	166	6.7	3,443	13.4
Self-pay or uninsured			63	21.7	0		346	19.8
Others			120	32.8	2	0.0	423	26.4

SOURCE: Test Site 1 and Test Site 2 data from October 1, 2013 to September 30, 2015. Test Site 3 from October 1, 2014 to September 30, 2015.

Performance by ICU exposure

We examined the rate at which antipsychotics are ordered in the ICU as compared to non-ICU settings. With the data available, we were able to look at this in two ways. Using method 1, we determined the unit where the patient was *first* assigned at the time of admission. If the patient was assigned to the ICU, the patient's encounter was classified as ICU. Other encounters were classified as non-ICU. Using method 2, we assigned patients to the ICU group if they were assigned to the ICU at *any* point in time during their encounter.

Both methods, as seen in Table 10, yield similar rates of antipsychotic ordering. Using method 1, 37.5 percent of ICU patients were ordered an antipsychotic during their encounter as compared to 21.4 percent of non-ICU patients. Using method 2, 37.7 percent of ICU encounters and 27.9 percent of non-ICU encounters had an antipsychotic order. Differences by ICU exposure were statistically significant, across both methods (X^2=87.7, p<.001).

Table 10. ICU versus non-ICU performance rates for antipsychotic ordering

	Method 1		Method 2	
	ICU	Non-ICU	ICU	Non-ICU
Performance rate	37.5	21.4	37.7	27.9

Performance by hospital

2019 resubmission: In Table 8a, we provide performance rates for each hospital across the for denominator conditions. A lower performance score indicates better care. Performance rates were higher in the second round of testing, across all denominator conditions. Across all conditions, Hospital 7 had the lowest performance scores. Hospital 7 also had the lowest performance score in the initial round of testing.

Table 8a. Comparison of performance rate based on exclusion criteria

	2019 Resubmission			
	Column A	Column B	Column C	Column D
	Original denominator (excluding encounters with Tourette's, Huntington's, schizophrenia, bipolar)	Original denominator plus exclusion of encounters with antipsychotics prior to admission	Original denominator plus exclusion of encounters with antipsychotics with a treatment-resistant depression diagnosis	Original denominator plus exclusion of encounters with antipsychotics prior to admission and/or antipsychotics with a treatment-resistant depression diagnosis
	Performance Rate %	Performance Rate %	Performance Rate %	Performance Rate %
Test Site 1				
Hospital 1	46.0%	38.7%	43.3%	36.0%
Hospital 2	37.7%	32.3%	35.3%	30.2%
Hospital 3	30.3%	21.7%	27.0%	19.2%
Hospital 4	35.4%	29.0%	32.2%	26.5%
Hospital 5	37.1%	31.1%	35.1%	29.7%
Hospital 6	35.0%	28.4%	31.3%	25.2%
Hospital 7	25.5%	20.5%	23.4%	19.0%
Hospital 8	48.1%	41.7%	45.2%	39.1%
Hospital 9	30.0%	23.2%	27.5%	21.2%

The variation in hospital-level measure performance is further illustrated in Figures 1a thru 1d, each representing a different denominator condition. The plots show the distribution of the performance rate and 95 percent confidence interval for each hospital relative to the overall performance rate across hospitals. The confidence interval is the range in which each hospital's performance rate would likely fall if extractions were repeated multiple times. Although some hospital rates were below the overall performance rate, across

denominator conditions, between two and four of the nine hospitals have measure rates significantly higher than the overall measure rate, indicating room for improvement.

Figure 1a. Distribution of performance rates by hospital – original measure specification

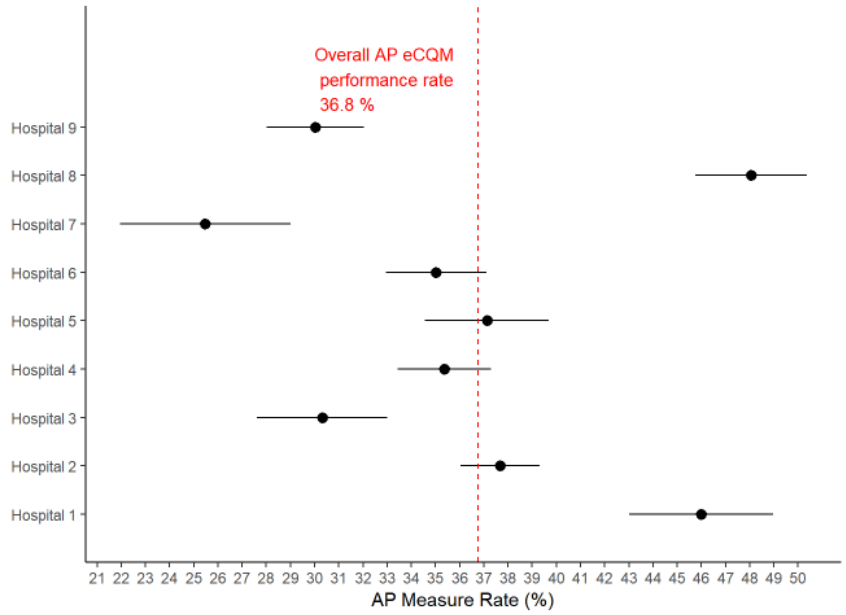


Figure 1b. Distribution of performance rates by hospital – with patients with antipsychotics prior to admission removed from the denominator

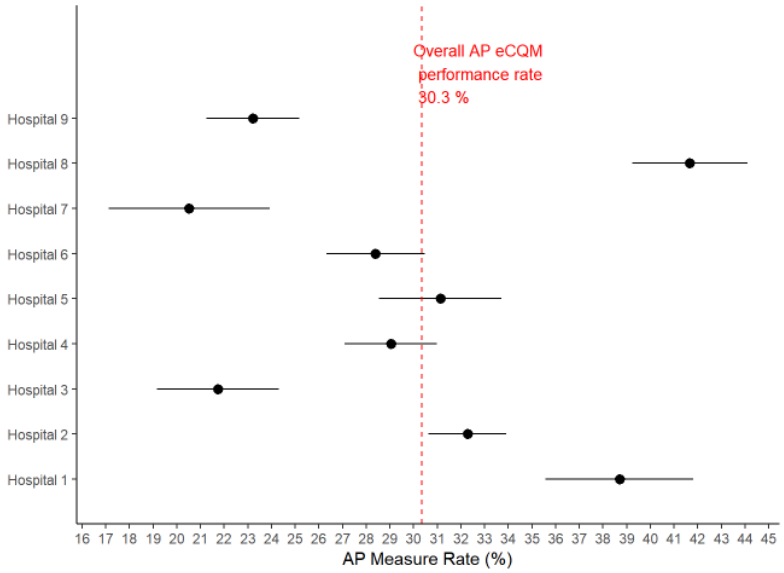


Figure 1c. Distribution of performance rates by hospital – with patients with antipsychotics and a treatment-resistant depression diagnosis removed from the denominator

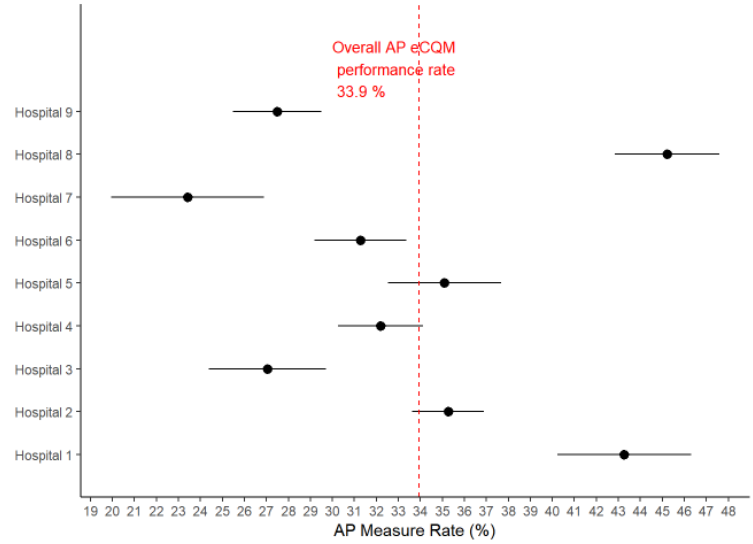
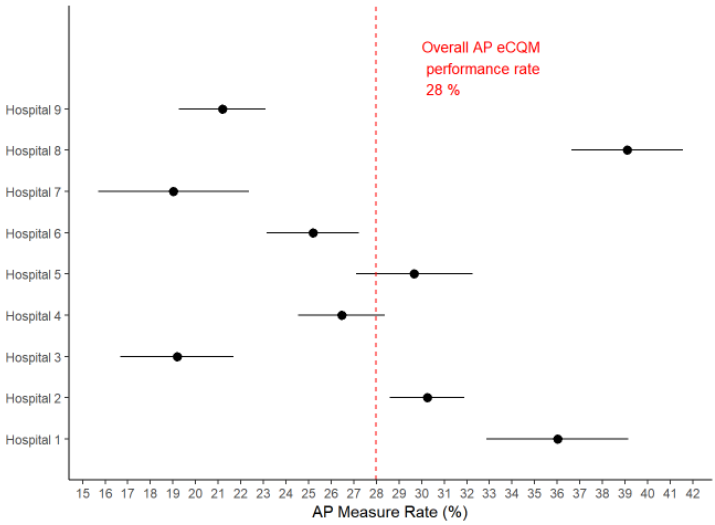


Figure 1d. Distribution of performance rates by hospital – with patients with antipsychotics prior to admission and/or patients with antipsychotics and a treatment resistant depression diagnosis removed from the denominator



Performance by disparity group

The disparities identified in the prior round of testing persisted in the more recent round of testing.

Age. Testing results indicated that antipsychotic use increases, across all denominator conditions, as age increases. As seen in Table 9a, among patients ages 65 to 74 years, across denominator conditions, the performance rate ranged from 21.1 to 28.8. Among patients 85 and older, the performance rate ranged from 41.0 to 50.5. This is an important finding as there has been significant concern about the inappropriate use of antipsychotics among older individuals. These findings support the notion that older patients are more likely to receive antipsychotics than younger patients, despite the AGS Beer's Criteria which cautions against their use in older adults. This lends support to the importance of this measure. Chi-squared analyses were done to exam subgroup differences. Across the three new denominator conditions, subgroup differences are statistically significant ($p<.001$) for age.

Sex, race, and ethnicity. As seen in Table 9a, males had higher rates of antipsychotic ordering than females across all denominator conditions. Across denominator conditions, performance rates among males ranged from 34.2 to 42.4. For females, the range was 23.4 to 32.6. With regard to race, Blacks were more likely than Whites to be ordered an antipsychotics, across all denominator conditions. Hispanic and non-Hispanic patients had similar performance rates across denominator conditions. Chi-squared analyses were done to exam subgroup differences. Across the three new denominator conditions, subgroup differences are statistically significant ($p<.001$) for sex and race. There were no significant difference in performance scores by ethnicity in any of the three denominator conditions.

Payer. Patients with Medicare, 92 percent of encounters) and Medicaid (1.3 percent of encounters) had the highest rate of antipsychotic ordering across denominator conditions. As expected, those with private insurance had the lowest performance rates. This is expected as most are younger and covered by insurance through their employers. Chi-squared analyses were done to exam subgroup differences. Across the three new denominator conditions, subgroup differences are statistically significant ($p<.001$) for payer.

Table 9a. Performance rate by patient encounter characteristic – original and 2019 resubmission

	2019 Resubmission							
	Original denominator (excluding encounters with Tourette's, Huntington's, schizophrenia, bipolar		Original denominator plus exclusion of encounters with antipsychotics prior to admission		Original denominator plus exclusion of encounters with antipsychotics with a treatment-resistant depression diagnosis		Original denominator plus exclusion of encounters with antipsychotics prior to admission and/or antipsychotics with a treatment-resistant depression diagnosis	
	n	Perf Rate* %	n	Perf Rate* %	n	Perf Rate* %	n	Perf Rate* %
Patient encounters	15697		14074		14980		13603	
Average age	77.3		77.3		78.2		77.3	
Age								
65 to 74	6720	28.8	6199	23.5	6435	25.8	6006	21.1
75 to 84	5435	37.6	4862	31.2	5184	34.8	4699	28.8
85 and older	3542	50.5	3013	43.2	3361	48.1	2898	41.0
Sex								
Male	6651	42.4	5893	35.9	6412	40.4	5736	34.2
Female	9042	32.6	8177	26.3	8564	29.1	7863	23.4
Unknown	4	50.0	4	50.0	4	50.0	4	50.0
Race								
White	7738	32.4	7025	26.2	7393	29.4	6806	23.9
Black	2007	40.8	1767	33.7	1912	38.1	1702	31.1
Other	5475	42.0	4844	35.5	5219	39.4	4671	33.2
Unknown	477	30.4	438	25.1	456	27.4	424	22.6
Ethnicity								
Hispanic	1528	37.0	1396	31.6	1463	34.4	1348	29.3
Non-Hispanic	13453	37.0	12014	30.4	12835	34.2	11613	28.0
Other or unknown								
Other	396	34.1	364	28.8	377	30.8	352	26.4
Unknown	320	28.4	300	24.0	305	24.9	290	21.4
(Primary) Payer								
Medicare	14441	37.2	12903	30.6	13774	34.3	12466	28.2
Medicaid	259	40.5	245	37.6	253	39.1	240	36.2
Private insurance	755	25.8	711	21.7	720	22.2	686	18.8
Self-pay/ uninsured	114	37.7	109	34.9	111	36.0	107	33.6
Others	128	46.1	106	36.8	122	43.4	104	35.6

SOURCE: Test Site 1, January 1, 2018 to December 31, 2018.

*Performance rate

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)
The results demonstrate that statistically significant differences can be detected between hospitals. The variations in performance across hospitals suggested meaningful differences in the quality of care provided between the lowest and highest performing hospitals and indicated that there is significant room for improvement. In addition, disparities in performance based on age, sex, race, ethnicity, and payer further suggested room for improvement. The statistically significant difference in antipsychotic ordering in the ICU versus non-ICU settings encouraged us to include stratification by unit of care in the measure specification.

2019 resubmission: Same as above with the exception that ICU status was not assessed during retesting.

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

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

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

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

Not applicable.

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Not applicable.

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
Date of birth is required for the measure calculation, as it is applicable for patients ages 65 years and older. In addition, encounters are defined by admission and discharge dates. Missing data on date of birth and admission and discharge dates was negligible. Missing data is not a threat to validity for this measure. The

majority of data elements required to calculate the performance rate are ones in which absence of data in a data field reflects the absence of a condition or behavior (for example, diagnosis or medication ordered).

2019 resubmission: Same as above.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)
See response for 2b6.1

2019 resubmission: Same as above.

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

See response for 2b6.1

2019 resubmission: Same as above.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic health records (EHRs)

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

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

Attachment: nqf_ecqm_feasibility_final_scorecard_AP_5_2019.xlsx

3c. Data Collection Strategy

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

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

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

Not applicable.

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

Value sets are housed in the Value Set Authority Center (VSAC), which is provided by the National Library of Medicine (NLM), in coordination with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

Viewing or downloading value sets requires a free Unified Medical Language System® (UMLS) Metathesaurus License, due to usage restrictions on some of the codes included in the value sets. Individuals interested in accessing value set content can request a UMLS license at (https://uts.nlm.nih.gov/license.html).

There are no other fees or licensing requirements to use this measure, which is in the public domain.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)

Public Reporting Payment Program Quality Improvement (external benchmarking to organizations) Quality Improvement (Internal to the specific organization)	

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

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

Not applicable; the measure is under initial endorsement review and is not currently used in an accountability program.

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

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

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

The measure has been submitted through the Measures Under Consideration process for the CMS Hospital Inpatient Quality Reporting Program and the Medicare and Medicaid Promoting Interoperability Program for Eligible Hospitals and Critical Access Hospitals (CAHs)

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

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

Not applicable

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

Not applicable

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

Describe how feedback was obtained.

Not applicable

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

Not applicable

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

Not applicable

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

Not applicable

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Adoption of this performance measure has the potential to improve the quality of care for hospitalized older adults in the area of patient safety, a priority area identified by the National Quality Strategy. Specifically, this measure will encourage thoughtful prescribing of antipsychotics for hospitalized patients and an increase in non-pharmacologic treatments and approaches to care. More careful antipsychotic prescribing among hospitalized individuals would be expected to result in fewer prescriptions continued after discharge and, ultimately, lower morbidity and mortality associated with the long-term use of these medications.

4b2. Unintended Consequences

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

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

This measure has not been implemented. During measure testing, experts suggested that a potential unintended consequence could be the increased use of alternative harmful medications such as benzodiazepines for delirium or behavioral and psychological

symptoms of dementia (BPSD). Additionally, clinical guidelines recognize that pharmacologic options should be a last resort after careful consideration and only after nonpharmacologic interventions have failed. This suggests that off-label antipsychotic use in the inpatient setting should not be expected to reach zero as clinical judgment will need to be exercised in situations where an alternative treatment may not be available or address the specific patient circumstances.

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

This measure has not been implemented. During measure testing, experts suggested that a potential benefit could be more thoughtful prescribing of antipsychotics in the inpatient setting, as well as fewer continued prescriptions after discharge to other care settings. This could encourage the use of delirium assessment and monitoring tools, improved detection of patient behaviors that could otherwise escalate to delirium, and the use of nonpharmacologic interventions to manage behavior.

5. Comparison to Related or Competing Measures

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

same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

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

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

The following are related measures currently endorsed by NQF:

NQF 2111: Antipsychotic Use in Persons with Dementia (Steward: PQA)

NQF 2993: Potentially Harmful Drug-Disease Interactions in the Elderly (DDE) (Steward: NCQA)

The following are related measures not currently endorsed by NQF:

CMS N011.01: Percentage of [Nursing Home] Residents Who Newly Received an Antipsychotic Medication (Short Stay) (Steward: CMS)

CMS N031.02: Percentage of [Nursing Home] Residents Who Received an Antipsychotic Medication (Long Stay) (Steward: CMS)

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

Yes

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

These measures are harmonized to the extent possible. While all measures assess the potentially inappropriate use of antipsychotic medications, this is the only measure that assesses use of antipsychotic medications in the inpatient hospital setting. CMS N011.01 and CMS N031.02 are intended for use in the nursing home setting. Measures NQF 2111 and NQF 2993 assess health plan performance. This measure's eligible population includes all patients in an inpatient hospital setting who are age 65 and older, which aligns with the age for measures NQF 2111 and NQF 2993. NQF 2111 and NQF 2993 only assess older adults with dementia, whereas this measure includes all older adults. The denominator exclusions are similar across measures. The exclusions in this measure—schizophrenia (including psychotic disorders), Tourette's syndrome, Huntington's disease, and bipolar disorder—are similar to exclusions in related measures. CMS N011.01, CMS N031.02, and NQF 2111 exclude patients with schizophrenia, Tourette's syndrome, or Huntington's disease. NQF 2111 also excludes patients with bipolar disorder. NQF 2993 excludes patients with psychosis, schizophrenia, or bipolar disorder. This measure also excludes from the numerator people in the inpatient setting who are identified as a threat to themselves or others. No other measure excludes these patients, although this exclusion is appropriate for the hospital setting. The specific antipsychotic medications included in each measure are aligned.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not applicable.

1. Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment Attachment: 1._Hospital-MDM_NQF_Form_Submission_v7.1_AP.docx

2. Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicare Services

Co.2 Point of Contact: Annese, Abdullah-Mclaughlin, annese.abdullah-mclaughlin@cms.hhs.gov, 410-786-2995-

Co.3 Measure Developer if different from Measure Steward: Mathematica, Inc.

Co.4 Point of Contact: Madeline, Pearse, name@mathematica-mpr.com%0b, 202-554-7564-

3. Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

Antipsychotics Measure Development Expert Work Group

This panel provided expertise in geriatric and inpatient care and provided feedback on the measure specifications and testing results.

--Byron Bair, MD, MBA - Salt Lake City VA

--Soo Borson, MD - University of Washington

--Josh Chodosh, MD MSHS - NYU School of Medicine

--Elizabeth Galik, RN, PhD, CRNP - University of Maryland School of Nursing

--Susan Merel, MD - University of Washington Department of Medicine

--Paul Rosenberg, MD - Johns Hopkins University

--Lynn Shell, PhD, APN, CARN-AP – Rutgers University

--Teepa Snow, MS, OTR/L, FAOTA - Positive Approach, LLC

--Heidi Wald, MD, MS, MS - University of Colorado

Technical Expert Panel

This panel provided overall guidance on measure development and project direction, including review of the measure specification and testing results.

--Peter Bach, MD, MAPP, Memorial Sloan Kettering

--James Burgess, PhD (co-chair) - Boston University

--Donna Slosburg, RN, BSN - ASC Quality Collaborative

--Ileana Pina, MD, MPH - Albert Einstein College of Medicine

--Jeremiah Schuur, MD, MHS - Brigham and Women's Hospital

--John Hertig, PharmD, MS - Purdue University

--Marc Overhage, PhD, MD - Siemens Health Services

--Kent Sepkowitz, MD - Memorial Sloan Kettering Cancer Center

--Maureen Dailey, PhD, RN - American Nurses Association

--Michael Howell, MD, MPH (chair) - University of Chicago Medicine

--Monica Peek, MD, MPH - Chicago Center for Diabetes Translation Research

--Nancy Foster - American Hospital Association

--Nathan Goldstein, MD - Mount Sinai School of Medicine

--Stephen Edge, MD - Baptist Cancer Center

--Susan McBride, PhD, RN-BC - Texas Tech University Health Sciences Center

--Thomas Louis, PhD - Johns Hopkins Bloomberg School of Public Health

Patient and Family Advisory Board

This panel provided feedback on the measure concept from the patient and family perspective.

--Darlene Barkman - Children's Hospital of Philadelphia

--Ann Cannarozzo - Rochester Regional Health System

--Maureen Corcoran - Cystic Fibrosis Foundation

--Ilene Corina - PULSE (Persons United Limiting Substandards and Errors in Healthcare) of NY

--John Harris - Johns Hopkins Hospital

--Toby Levin - Suburban Hospital Patient and Family Advisory Council

--Christopher Mason - Peace Health Patient Advisory Council

--Teresa Masters - Patient and Family Centered Council, University of California, San Diego

--Lisa McDermott - National Brain Tumor Society

--Kelly Parent - Patient and Family Centered Care Program, University of Michigan Health System

--Lee Tomlinson - Center for More Compassionate Care

--Karel Shapiro - Rochester General Hospital

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

Ad.4 What is your frequency for review/update of this measure? Specifications for this eCQM will be reviewed and updated annually.

Ad.5 When is the next scheduled review/update for this measure? 05, 2020

Ad.6 Copyright statement: Limited proprietary coding is contained in the Measure specifications for user convenience. Users of proprietary code sets should obtain all necessary licenses from the owners of the code sets.

CPT(R) contained in the Measure specifications is copyright 2004-2016 American Medical Association. LOINC(R) copyright 2004-2016 Regenstrief Institute, Inc. This material contains SNOMED Clinical Terms(R) (SNOMED CT[R]) copyright 2004-2016 International Health Terminology Standards Development Organisation. ICD-10 copyright 2016 World Health Organization. All Rights Reserved.

Ad.7 Disclaimers: These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. The measures and specifications are provided without warranty.

Ad.8 Additional Information/Comments: Not applicable