



## Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

### Brief Measure Information

**NQF #:** 3747

**Corresponding Measures:**

**Measure Title:** Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization

**Measure Steward:** New York State Office of Mental Health

**sp.02. Brief Description of Measure:** The percentage of discharges for members 6 years of age and older who were hospitalized for treatment of selected mental illness or intentional self-harm diagnoses and who had at least five follow-up community-based mental health care visits in the 90 days after discharge.

**1b.01. Developer Rationale:**

---

**sp.12. Numerator Statement:** To meet the numerator criteria for this measure, discharges must receive five or more follow-up visits with a community-based mental health care provider within 90 days after discharge for inpatient treatment of select mental health or intentional self-harm diagnoses. Follow-up visits that occur on the date of discharge are not included.

**sp.14. Denominator Statement:** The eligible population for this measure is acute inpatient discharges ages 6-64 principally hospitalized for select mental illnesses or intentional self-harm and enrolled in Medicaid on the date of discharge through 90 days after discharge. Discharges with acute direct transfers, acute readmissions with total length of stay of 42 or more days, non-acute direct transfers, and non-acute readmissions are excluded. If members have more than one discharge during the measurement year, all discharges during the measurement year will be included.

**sp.16. Denominator Exclusions:** In addition to the discharges with acute direct transfers, certain acute readmissions, non-acute direct transfers, and non-acute readmissions detailed above, discharges who are dually enrolled in Medicare and Medicaid and discharges in hospice or using hospice services anytime during the measurement year are excluded.

---

**Measure Type:** Process

**sp.28. Data Source:**

Claims

**sp.07. Level of Analysis:**

Health Plan

---

**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:**

**sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:**

## 1. Importance to Measure and Report

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

---

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

**Current Submission:**

Updated evidence information here.

**Previous (Year) Submission:**

Evidence from the previous submission here.

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

Group 1 - Evidence - Systematic Reviews Table

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

**1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.**

[Response Begins]

[Response Ends]

**1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.**

*Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., “topped out”, disparities data may demonstrate an*

*opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

## 2. Scientific Acceptability of Measure Properties

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

---

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

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

**[Response Begins]**

Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization

**[Response Ends]**

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

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

**[Response Begins]**

The percentage of discharges for members 6 years of age and older who were hospitalized for treatment of selected mental illness or intentional self-harm diagnoses and who had at least five follow-up community-based mental health care visits in the 90 days after discharge.

**[Response Ends]**

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

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

Please do not select:

- Surgery: General

**[Response Begins]**

Behavioral Health

Behavioral Health: Anxiety

Behavioral Health: Bipolar Disorder

Behavioral Health: Depression

Behavioral Health: Other Serious Mental Illness

Behavioral Health: Post-Traumatic Stress Disorder (PTSD)

Behavioral Health: Schizophrenia

Behavioral Health: Suicide

**[Response Ends]**

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

**[Response Begins]**

Access to Care

Care Coordination

Care Coordination: Transitions of Care

**[Response Ends]**

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

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

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

*Please do not select:*

- *Populations at Risk: Populations at Risk*

**[Response Begins]**

Adults (Age >= 18)

Children (Age < 18)

**[Response Ends]**

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

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

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

*Please do not select:*

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

**[Response Begins]**

Health Plan

**[Response Ends]**

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

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

**[Response Begins]**

Behavioral Health

Post-Acute Care

**[Response Ends]**

**sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.**

*Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".*



**[Response Begins]**

None available.

**[Response Ends]**

**sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.**

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

**[Response Begins]**

Available in attached Excel or csv file

**[Response Ends]**

Attachment: 3747\_EIC Discharge Diagnosis Value Set.xlsx

**sp.13. State the numerator.**

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

*DO NOT include the rationale for the measure.*

**[Response Begins]**

To meet the numerator criteria for this measure, discharges must receive five or more follow-up visits with a community-based mental health care provider within 90 days after discharge for inpatient treatment of select mental health or intentional self-harm diagnoses. Follow-up visits that occur on the date of discharge are not included.

**[Response Ends]**

**sp.14. Provide details needed to calculate the numerator.**

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

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

**[Response Begins]**

Five of any of the following within 90 days of discharge meet the criteria for community-based mental health follow-up visits.

- An outpatient visit (Visit Setting Unspecified Value Set) with (Outpatient POS Value Set) with a mental health provider.
- An outpatient visit (BH Outpatient Value Set) with a mental health provider.
- An intensive outpatient encounter or partial hospitalization (Visit Setting Unspecified Value Set) with (Partial Hospitalization POS Value Set).

- An intensive outpatient encounter or partial hospitalization (Partial Hospitalization or Intensive Outpatient Value Set).
- A community mental health center visit (Visit Setting Unspecified Value Set; BH Outpatient Value Set; Observation Value Set; Transitional Care Management Services Value Set) with (Community Mental Health Center POS Value Set).
- Electroconvulsive therapy (Electroconvulsive Therapy Value Set) with (Ambulatory Surgical Center POS Value Set; Community Mental Health Center POS Value Set; Outpatient POS Value Set; Partial Hospitalization POS Value Set).
- A telehealth visit: (Visit Setting Unspecified Value Set) with (Telehealth POS Value Set) with a mental health provider.
- An observation visit (Observation Value Set) with a mental health provider.
- Transitional care management services (Transitional Care Management Services Value Set), with a mental health provider.
- A visit in a behavioral healthcare setting (Behavioral Healthcare Setting Value Set).
- A telephone visit (Telephone Visits Value Set) with a mental health provider.
- Psychiatric collaborative care management (Psychiatric Collaborative Care Management Value Set).

**[Response Ends]**

**sp.15. State the denominator.**

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

**[Response Begins]**

The eligible population for this measure is acute inpatient discharges ages 6-64 principally hospitalized for select mental illnesses or intentional self-harm and enrolled in Medicaid on the date of discharge through 90 days after discharge. Discharges with acute direct transfers, acute readmissions with total length of stay of 42 or more days, non-acute direct transfers, and non-acute readmissions are excluded. If members have more than one discharge during the measurement year, all discharges during the measurement year will be included.

**[Response Ends]**

**sp.16. Provide details needed to calculate the denominator.**

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

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

**[Response Begins]**

<b>Product lines</b>	Medicaid
<b>Ages</b>	6 – 64 years old as of the date of discharge. Report two age stratifications and a total rate: <ol style="list-style-type: none"> <li>1. 6-20 years old</li> <li>2. 21-64 years old</li> <li>3. Total</li> </ol>
<b>Continuous Enrollment</b> <b>Allowable gap</b>	Date of discharge through 90 days after discharge No gaps in enrollment
<b>Anchor date</b>	None.
<b>Benefits</b>	Medical, Mental Health, and Substance Use.
<b>Event/diagnosis</b>	<p>An acute inpatient discharge with a principal diagnosis of mental illness or intentional self-harm (<u>Mental Health Discharge Diagnosis Value Set</u>; <u>Intentional Self-Harm Value Set</u>) on the discharge claim on or between October 1 of the year prior to the measurement year and September 30 of the measurement year. To identify acute inpatient discharges:</p> <ol style="list-style-type: none"> <li>1. Identify all acute and nonacute inpatient stays (<u>Inpatient Stay Value Set</u>).</li> <li>2. Exclude nonacute inpatient stays (<u>Nonacute Inpatient Stay Value Set</u>).</li> <li>3. Identify the discharge date for the stay.</li> </ol> <p>The denominator for this measure is based on discharges, not on members. If members have more than one discharge, include all discharges on or between October 1 of the year prior to the measurement year and September 30 of the measurement year.</p>

<b>Product lines</b>	Medicaid
<b>Acute readmission or direct transfers</b>	Follow the steps below to exclude discharges with acute direct transfers, certain acute readmissions, non-acute direct transfers, and non-acute readmissions.
<b>Nonacute readmission or direct transfer</b>	<p>Identify readmissions and direct transfers to an acute inpatient care setting during the 90-day follow-up period:</p> <ol style="list-style-type: none"> <li>1. Identify all acute and nonacute inpatient stays (<a href="#">Inpatient Stay Value Set</a>).</li> <li>2. Exclude nonacute inpatient stays (<a href="#">Nonacute Inpatient Stay Value Set</a>).</li> <li>3. Identify the admission date for the stay (the admission date must occur during the 90 day follow up period). If the admission date is the same date or one day later than the initial discharge, the admission is considered a direct transfer.</li> <li>4. Identify the discharge date for the stay</li> <li>5. For readmissions (admission dates 2-90 days after the initial discharge), calculate the length of stay in days (discharge date of the stay minus the admission date of the stay).</li> </ol> <p>Exclude both the initial discharge and direct transfer discharge if the last discharge occurs after September 30 of the measurement year.</p> <p>If the direct transfer to the acute inpatient care setting was for a principal diagnosis (use only the principal diagnosis on the discharge claim) of mental health disorder or intentional self-harm (Mental Health Diagnosis Value Set; Intentional Self-Harm Value Set), exclude the initial discharge.</p> <p>If the direct transfer to the acute inpatient care setting was for any other principal diagnosis (use only the principal diagnosis on the discharge claim) exclude both the original and the direct transfer discharge.</p> <p>Calculate the total length of stay in days for acute readmissions 2-90 days after the initial discharge. If this sum is 42 days or more, exclude the initial discharge.</p> <p>Exclude discharges followed by readmission or direct transfer to a nonacute inpatient care setting within the 90-day follow-up period, regardless of principal diagnosis for the readmission. To identify readmissions and direct transfers to a nonacute inpatient care setting:</p> <ol style="list-style-type: none"> <li>1. Identify all acute and nonacute inpatient stays (<a href="#">Inpatient Stay Value Set</a>).</li> <li>2. Confirm the stay was for nonacute care based on the presence of a nonacute code (<a href="#">Nonacute Inpatient Stay Value Set</a>) on the claim.</li> <li>3. Identify the admission date for the stay.</li> </ol> <p>These discharges are excluded from the measure because rehospitalization or direct transfer may prevent an outpatient follow-up visit from taking place.</p>

[Response Ends]

**sp.17. Describe the denominator exclusions.**

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

[Response Begins]

In addition to the discharges with acute direct transfers, certain acute readmissions, non-acute direct transfers, and non-acute readmissions detailed above, discharges who are dually enrolled in Medicare and Medicaid and discharges in hospice or using hospice services anytime during the measurement year are excluded.

**[Response Ends]**

**sp.18. Provide details needed to calculate the denominator exclusions.**

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

**[Response Begins]**

To identify Medicare and Medicaid dual enrollment, Medicare enrollment data should be queried. If the discharge was enrolled in Medicare any time between the date of discharge through 90 days post discharge, exclude the discharge from the denominator.

Exclude discharges who use hospice services or elect to use a hospice benefit any time during the measurement year, regardless of when the services began. These discharges may be identified using the NCQA HEDIS Hospice Encounter Value Set or NCQA HEDIS Hospice Intervention Value Set.

**[Response Ends]**

**sp.19. Provide all information required to stratify the measure results, if necessary.**

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

**[Response Begins]**

Two age stratifications and a total rate should be reported. Age should be calculated at the date of discharge as date of discharge – date of birth:

- Stratification 1: 6-20 years old
- Stratification 2: 21-64 years old
- Total Rate

**[Response Ends]**

**sp.20. Is this measure adjusted for socioeconomic status (SES)?**

**[Response Begins]**

No

**[Response Ends]**

**sp.21. Select the risk adjustment type.**

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

**[Response Begins]**

No risk adjustment or risk stratification

**[Response Ends]**

**sp.22. Select the most relevant type of score.**

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

**[Response Begins]**

Rate/proportion

**[Response Ends]**

Attachment: 3747\_Engagement in Care Sample Report.xlsx

**sp.23. Select the appropriate interpretation of the measure score.**

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

**[Response Begins]**

Better quality = Higher score

**[Response Ends]**

**sp.24. Diagram or describe the calculation of the measure score as an ordered sequence of steps.**

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

**[Response Begins]**

**Step One – calculate the eligible population as detailed below:**

Product lines	Medicaid
Ages	6 – 64 years old as of the date of discharge. Report two age stratifications and a total rate:  1. 6-20 years old 2. 21-64 years old 3. Total
Continuous Enrollment	Date of discharge through 90 days after discharge
Allowable gap	No gaps in enrollment.
Anchor date	None.
Benefits	Medical, Mental Health, and Substance Use.

Product lines	Medicaid
Event/diagnosis	<p>An acute inpatient discharge with a principal diagnosis of mental illness or intentional self-harm (<u>HEDIS Mental Illness Value Set with NYS enhancements</u>; <u>Intentional Self-Harm Value Set</u>) on the discharge claim on or between October 1 of the year prior to the measurement year and September 30 of the measurement year. To identify acute inpatient discharges:</p> <ol style="list-style-type: none"> <li>1. Identify all acute and nonacute inpatient stays (<u>Inpatient Stay Value Set</u>).</li> <li>2. Exclude nonacute inpatient stays (<u>Nonacute Inpatient Stay Value Set</u>).</li> <li>3. Identify the discharge date for the stay.</li> </ol> <p>The denominator for this measure is based on discharges, not on members. If members have more than one discharge, include all discharges on or between October 1 of the year prior to the measurement year and September 30 of the measurement year.</p>

Product lines	Medicaid
Acute readmission or direct transfers	Follow the steps below to exclude discharges with acute direct transfers, certain acute readmissions, non-acute direct transfers, and non-acute readmissions.
Nonacute readmission or direct transfer	Identify readmissions and direct transfers to an acute inpatient care setting during the 90-day follow-up period:
Exclusions	<ol style="list-style-type: none"> <li>1. Identify all acute and nonacute inpatient stays (<a href="#">Inpatient Stay Value Set</a>).</li> <li>2. Exclude nonacute inpatient stays (<a href="#">Nonacute Inpatient Stay Value Set</a>).</li> <li>3. Identify the admission date for the stay (the admission date must occur during the 90 day follow-up period). If the admission date is the same date or one day later than the initial discharge, the admission is considered a direct transfer.</li> <li>4. Identify the discharge date for the stay</li> <li>5. For readmissions (admission dates 2-90 days after the initial discharge), calculate the length of stay in days (discharge date of the stay minus the admission date of the stay).</li> <li>6. Exclude both the initial discharge and direct transfer discharge if the last discharge occurs after September 30 of the measurement year.</li> <li>7.</li> <li>8. If the direct transfer to the acute inpatient care setting was for a principal diagnosis (use only the principal diagnosis on the discharge claim) of mental health disorder or intentional self-harm (Mental Health Diagnosis Value Set; Intentional Self-Harm Value Set), exclude the initial discharge.</li> <li>9.</li> <li>10. If the direct transfer to the acute inpatient care setting was for any other principal diagnosis (use only the principal diagnosis on the discharge claim) exclude both the original and the direct transfer discharge.</li> </ol> <p>Calculate the total length of stay in days for acute readmissions 2-90 days after the initial discharge. If this sum is 42 days or more, exclude the initial discharge.</p> <p>Exclude discharges followed by readmission or direct transfer to a nonacute inpatient care setting within the 90-day follow-up period, regardless of principal diagnosis for the readmission. To identify readmissions and direct transfers to a nonacute inpatient care setting:</p> <ol style="list-style-type: none"> <li>1. Identify all acute and nonacute inpatient stays (<a href="#">Inpatient Stay Value Set</a>).</li> <li>2. Confirm the stay was for nonacute care based on the presence of a nonacute code (<a href="#">Nonacute Inpatient Stay Value Set</a>) on the claim.</li> <li>3. Identify the admission date for the stay.</li> </ol> <p>These discharges are excluded from the measure because rehospitalization or direct transfer may prevent an outpatient follow-up visit from taking place.</p> <p>Medicare duals are excluded. Discharges in hospice or using hospice services anytime during the measurement year are excluded.</p>

**Step Two – include all discharges in the eligible population in the measure denominator.**

**Step Three – determine how many of the discharges in the denominator meet the numerator criteria as detailed below:**



1. To be included in the numerator, a discharge must have received five or more follow-up visits with a community-based mental health care provider within 90 days after discharge. Do not include visits that occur on the date of discharge. Any of the following meet the criteria for a follow-up visit.
  - An outpatient visit (Visit Setting Unspecified Value Set) with (Outpatient POS Value Set) with a mental health provider.
  - An outpatient visit (BH Outpatient Value Set) with a mental health provider.
  - An intensive outpatient encounter or partial hospitalization (Visit Setting Unspecified Value Set) with (Partial Hospitalization POS Value Set).
  - An intensive outpatient encounter or partial hospitalization (Partial Hospitalization or Intensive Outpatient Value Set).
  - A community mental health center visit (Visit Setting Unspecified Value Set; BH Outpatient Value Set; Observation Value Set; Transitional Care Management Services Value Set) with (Community Mental Health Center POS Value Set).
  - Electroconvulsive therapy (Electroconvulsive Therapy Value Set) with (Ambulatory Surgical Center POS Value Set; Community Mental Health Center POS Value Set; Outpatient POS Value Set; Partial Hospitalization POS Value Set).
  - A telehealth visit: (Visit Setting Unspecified Value Set) with (Telehealth POS Value Set) with a mental health provider.
  - An observation visit (Observation Value Set) with a mental health provider.
  - Transitional care management services (Transitional Care Management Services Value Set), with a mental health provider.
  - A visit in a behavioral healthcare setting (Behavioral Healthcare Setting Value Set).
  - A telephone visit (Telephone Visits Value Set) with a mental health provider.
  - Psychiatric collaborative care management (Psychiatric Collaborative Care Management Value Set).

**[Response Ends]**

**sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.**

*Examples of samples used for testing:*

- *Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.*
- *The sample should represent the variety of entities whose performance will be measured. The [2010 Measure Testing Task Force](#) recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.*
- *The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.*
- *When possible, units of measurement and patients within units should be randomly selected.*

**[Response Begins]**

Not applicable. Testing for this measure was not based on a sample.

**[Response Ends]**

**sp.30. Select only the data sources for which the measure is specified.**

**[Response Begins]**

Claims

**[Response Ends]**

**sp.31. Identify the specific data source or data collection instrument.**

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

**[Response Begins]**

Medicaid administrative data.

**[Response Ends]**

**sp.32. Provide the data collection instrument.**

**[Response Begins]**

No data collection instrument provided

**[Response Ends]**

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

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

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

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

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

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

AND

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

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

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

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

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

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

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

## Definitions

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

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by

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

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

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

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

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

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

**Current Submission:**

Updated testing information here.

**Previous (Year) Submission:**

Testing from the previous submission here.

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

**[Response Begins]**

Claims

**[Response Ends]**

**2a.02. If an existing dataset was used, identify the specific dataset.**

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

**[Response Begins]**

New York State administrative Medicaid data.

**[Response Ends]**

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

*Use the following format: "MM-DD-YYYY - MM-DD-YYYY"*

**[Response Begins]**

1-1-2018 – 12-31-2018

**[Response Ends]**

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

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

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

*Please do not select:*

- Clinician: Clinician
- Population: Population

**[Response Begins]**

Health Plan

**[Response Ends]**

**2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).**

*Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.*

**[Response Begins]**

Thirty-one Medicaid managed care plans that operate in New York State were included in the analysis. Medicaid fee-for-service discharges were also included as a separate entity.

**[Response Ends]**

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

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

**[Response Begins]**

50,234 discharges were included in the analysis. A sample was not used; this number represents all Calendar Year 2018 discharges in the measure eligible population.

Descriptive characteristics of the population included in the analyses are below.

Table 1. Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization, Age at Discharge of the Eligible Population, 2018 NYS Medicaid Discharges

Age Group	Frequency	Percent
<10	1018	2.03
10-19	10742	21.38
20-29	13434	26.74
30-39	10371	20.65

Age Group	Frequency	Percent
40-49	6784	13.50
50-59	6402	12.74
60-64	1483	2.95

Table 2. Sex of the Eligible Population, 2018 NYS Medicaid Discharges

Sex	Frequency	Percent
Female	23203	46.19
Male	27031	53.81

Table 3. Ethnicity of the Eligible Population, 2018 NYS Medicaid Discharges

Ethnicity	Frequency	Percent
Hispanic or Latino	9186	18.29
Non - Hispanic or Latino	37839	75.33
Unknown	3209	6.39

Table 4. Race of the Eligible Population, 2018 NYS Medicaid Discharges

Race Group	Frequency	Percent
American Indian/Alaska Native	506	1.01
Asian	2483	4.94
Black/African American	17444	34.73
Native Hawaiian/Other Pacific Islander	272	0.54
Unknown	3647	7.26
White	23660	47.10
Two or More Races	2222	4.42

Table 5. Primary Diagnosis at Discharge of the Eligible Population, 2018 NYS Medicaid Discharges

Primary Diagnosis Category at Discharge	Frequency	Percent
Anxiety Disorders	7	0.01
Bipolar and Related Disorders	9095	18.11
Depressive Disorders	16035	31.92
Disruptive, Impulse-Control, and Conduct Disorders	887	1.77
Dissociative Disorders	11	0.02
Feeding and Eating Disorders	3	0.01
Intentional Self-harm	903	1.80
Medication-Induced Movement Disorders	1	0.00
Neurodevelopmental Disorders	802	1.60
Obsessive-Compulsive and Related Disorders	51	0.10
Other Mental Disorders	1	0.00
Personality Disorders	605	1.20
Schizophrenia Spectrum and Other Psychotic Disorders	19618	39.05
Somatic Symptom and Related Disorders	16	0.03

Primary Diagnosis Category at Discharge	Frequency	Percent
Substance-Related and Addictive Disorders	7	0.01
Trauma- and Stressor-Related Disorders	2192	4.36

[Response Ends]

**2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.**

[Response Begins]

The data do not differ between different aspects of testing.

[Response Ends]

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

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

[Response Begins]

Information on social risk factors is not available in NYS Medicaid administrative data. The NYS Office of Mental Health attempted to collect social risk factor information through an assessment for Health and Recovery Plan (HARP) members (a specialty Medicaid product line for individuals with serious behavioral health conditions), but the assessment was completed for less than 20% of enrolled individuals annually. We do not feel that this data is representative enough of the measure eligible population to use in measure testing. NYS OMH is exploring other means of collecting social determinants of health information in the future, including an annual assessment for patients of NYS mental health outpatient clinics and/or through EHRs, so that this important data can be incorporated into performance measurement development.

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter “see validity testing section of data elements”; and enter “N/A” for 2a.11 and 2a.12.

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

*Choose one or both levels.*

[Response Begins]

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

[Response Ends]

**2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.**

*Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.*

**[Response Begins]**

To test reliability, we performed signal to noise analysis using beta binomial models. Specifically, we used the methodology from “The Reliability of Provider Profiling, A Tutorial” by John L. Adams for the National Committee for Quality Assurance, RAND Health, 2009. This method tests how well the performance of one reporting entity can be distinguished from another. Conceptually, it is the ratio of signal to noise and is appropriate for testing the reliability of measures with a binary (yes/no, pass/fail, etc.) outcome. The signal is the proportion of variation attributable to plan performance and the noise is the proportion of variation attributable to measurement error.

A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance.

Our steps to calculate reliability are shown below. Health plans were used as the reporting entity.

The formula for signal-to-noise reliability is:

Signal-to-noise reliability =  $\sigma^2 \text{plan-to-plan} / (\sigma^2 \text{plan-to-plan} + \sigma^2 \text{error})$

We estimated the variance between plans and variance within plans. The formulas for the variance calculations are:

1. Variance between plans =  $\sigma^2 \text{plan-to-plan} = (\alpha \beta) / (\alpha + \beta + 1)(\alpha + \beta)^2$

$\alpha$  and  $\beta$  are two shape parameters of the Beta-Binomial distribution,  $\alpha > 0$ ,  $\beta > 0$

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

$\hat{p}$  = observed rate for the plan

$n$  = plan-specific denominator for the observed rate

We ran a SAS Macro program that was developed by John Adams and based on his tutorial mentioned above to calculate the reliability for each reporting entity. To derive the mean signal-to-noise reliability presented in Table 6, the average of the entity level reliability estimates was taken. The mean signal-to-noise reliability indicates the measure’s mean ability to differentiate between reporting entity performance.

Table 7 shows the standard error (SE) and 95% confidence interval (95% CI) of the mean signal-to-noise reliability for all reporting entities. The reporting entities were divided into terciles based on the denominator size (number of eligible discharges per plan). The SE and 95% CI of the mean signal-to-noise reliability provides information about the stability of the reliability estimates. The formula for the 95% CIs is the mean signal-to-noise reliability  $\pm$  (1.96\*SE).

Table 8 shows the minimum, maximum, 10th, 25th, 50th, 75th, and 90<sup>th</sup> percentiles of the plan-level signal-to-noise reliability estimates. The reporting entities were divided into terciles based on the denominator size (number of eligible discharges per plan). Each plan’s reliability estimate was calculated as described above.

**[Response Ends]**

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

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

**[Response Begins]**

Table 6. Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization, Point Estimate of Mean Signal-to-Noise Reliability, 2018 NYS Medicaid Discharges



Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization	Mean Signal-to-Noise Reliability
All NYS Medicaid Eligible Discharges	0.946

Table 7. Mean Signal-To-Noise Reliability, Standard Error (SE) and 95% Confidence Interval (95% CI) by Terciles of the Denominator Size, 2018 NYS Medicaid Discharges

Tercile	Number of Plans	Number of Eligible Discharges per Plan (min - max)	Mean Signal-To-Noise Reliability	SE	95% CI
All NYS Medicaid Eligible Discharges	32	57-12467	0.946	0.011	(0.924, 0.967)
Tercile 1	10	57-237	0.870	0.019	(0.833, 0.906)
Tercile 2	11	298-1026	0.969	0.004	(0.960, 0.977)
Tercile 3	11	1178-12467	0.992	0.002	(0.989, 0.995)

Table 8. Distribution of Plan-Level Signal-To-Noise Reliability by Terciles of the Denominator Size, 2018 NYS Medicaid Discharges

		Distribution of Plan Estimates of Signal-to-Noise Reliability						
Tercile	Number of Plans	Min	P10	P25	P50	P75	P90	Max
All NYS Medicaid Eligible Discharges	32	0.756	0.853	0.926	0.976	0.987	0.996	0.999
Tercile 1	10	0.756	0.759	0.836	0.881	0.926	0.928	0.928
Tercile 2	11	0.942	0.942	0.955	0.975	0.980	0.982	0.982
Tercile 3	11	0.985	0.985	0.987	0.992	0.996	0.998	0.999

[Response Ends]

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

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

[Response Begins]

A reliability of 0.70 is often considered adequate, 0.80 is generally described as good, and a reliability of greater than 0.90 is thought to be high.

The mean reliability estimate for NYS Medicaid plans was 0.946 with a 95% CI of 0.924 to 0.967, which falls into the high reliability category (Table 1). Across the terciles of plans, mean reliability estimates ranged from 0.870, 95% CI (0.833-0.906) to 0.992, 95% CI (0.989-0.995), which indicate good to high reliability (Table 2). Plan level reliability increased as denominator size increased. The lowest reliability estimate among all NYS Medicaid plans was 0.756, which is considered adequate. All plans in the top two terciles of denominator size had reliability estimates above 0.90 and fell into the high reliability category.

**[Response Ends]**

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

**[Response Begins]**

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

Empirical validity testing

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)

**[Response Ends]**

**2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.**

*Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.*

**[Response Begins]**

Construct validity, empirical validity, and face validity were tested for this measure.

To test construct validity, we calculated the Pearson correlation coefficient between Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization and the NCQA HEDIS measure Follow-Up After Hospitalization for Mental Illness. We expected to find a moderate positive correlation between the measures. The eligible population for these measures is similar and there is overlap between the group meeting the criteria for Follow-Up After Hospitalization for Mental Illness (one visit 7 or 30 days after discharge) and meeting the criteria for Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization. In our Calendar Year 2018 discharge cohort, 62% of discharges meeting the criteria for the 30-day Follow-Up After Hospitalization for Mental Illness had five community based mental health follow-up visits in the 90 days after discharge. The two measures are conceptually different, because the HEDIS measure is one of short-term follow-up and the NYS OMH measure is one of longer-term engagement in care. A moderate correlation is desirable in this situation because the two measures are similar but not identical.

The Pearson correlation measures the strength and direction of the linear relationship between two continuous variables. Correlation coefficients range from 1 to -1. A value of 1 represents a perfect positive correlation between two variable and a value of -1 represents a perfect negative correlation. A value of zero means there is no linear relationship between the variables.

To test empirical validity, we calculated Concordance Statistics (or C Statistics) between Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization and three outcomes: mental health inpatient readmissions, psychotropic medication adherence, and continued engagement in care at six months post discharge. A mental health inpatient readmission was defined as an acute admission with a primary mental health or intentional self-harm diagnosis in months four through nine after discharge. Medication adherence was defined as having 80% or more days covered with a medication class appropriate for the primary diagnosis in months four through six following discharge. Continued engagement in care at six months post discharge was defined as having at least one community based mental health care visit in month six post discharge.

The predictive accuracy of a statistical model can be measured by the agreement between observed and predicted outcomes. The C (Concordance) Statistic assesses the ability of a risk factor to predict an outcome. It is commonly used with logistic regression models with a binary outcome. The concept underlying concordance is that a subject who experiences a particular outcome has a higher predicted probability of that outcome than a subject who does not experience the outcome. The C Statistic can be calculated as the proportion of pairs of subjects whose observed and predicted outcomes agree (are concordant) among all possible pairs in which one subject

experiences the outcome of interest and the other one does not. The higher the C-statistic, the better the model can discriminate between subjects who do experience the outcome of interest and subjects who do not.

C statistics generally range from 0.5 to 1. A C statistic of 1 represents perfect concordance and 0.5 means the model is no better at predicting the outcome than random chance. Models are typically considered reasonable when the C statistic is higher than 0.7 and strong when it exceeds 0.8.

To ensure face validity, this measure was conceptualized by a workgroup consisting of mental health clinicians and researchers. The concept of follow-up care after a psychiatric discharge being beneficial was supported by a literature review and the use of the HEDIS Follow-Up After Hospitalization for Mental Illness measure by NYS Medicaid to monitor quality of care. The workgroup and NYS OMH program staff agreed that measuring engagement beyond a single follow-up visit was clinically sound. They felt that more than a single visit was needed to adequately treat an individual after a discharge.

During the development process, we presented the measure concepts and testing analyses to the NYS Behavioral Health Clinical Advisory group and gathered their feedback. This group consists of representatives from NYS state health and behavioral health agencies, managed care plans, providers, and behavioral health advocacy organizations.

**[Response Ends]**

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

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

**[Response Begins]**

Table 9. Pearson Correlation Coefficient for New York State Medicaid Health Plans for Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization and Follow-Up After Hospitalization for Mental Illness, 2018 NYS Medicaid Discharges

	Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization
HEDIS Follow-Up After Hospitalization for Mental Illness	0.56044 N=50,234 discharges over 31 plans and FFS, p value <.0001

Table 10. Concordance Statistics for Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization, NYS Medicaid Calendar Year 2018 Discharges

Outcome	Population	C Statistic
Inpatient mental health readmissions in months 4-9 after discharge	All eligible discharges	0.5291
Adherence to antipsychotic medications in months 4-6 after discharge	Discharges with a primary diagnosis of schizophrenia	0.6428
Adherence to mood stabilizer medications in months 4-6 after discharge	Discharges with a primary diagnosis of bipolar disorder	0.6461
Adherence to antidepressant medications in months 4-6 after discharge	Discharges with a primary diagnosis of depression	0.6367

Outcome	Population	C Statistic
At least one community based mental health visit in month six post discharge	All eligible discharges	0.7244

[Response Ends]

**2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)**

[Response Begins]

For construct validity, the correlation between our measure and the HEDIS FUH measure is moderate. This result is expected because 62% of the 2018 discharges that met the criteria for the HEDIS measure met the criteria for the NYS OMH measure. The strength of the correlation is reasonable when considering the degree of overlap between the two measures.

In terms of empirical validity, our measure was reasonably predictive of engagement in community based mental health care at six months post discharge. The measure did not meet the 0.7 C Statistic threshold for predictability for acute service use or medication adherence. This lack of predictability may be due to confounding as we did not adjust for any risk factors in our models.

For face validity the measure specifications, descriptive analyses, reliability analyses, and validity analyses were presented to internal workgroups and the NYS Behavioral Health Clinical Advisory group over a several month period. The workgroups agreed that the measure had good face validity.

[Response Ends]

**2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.**

*Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.*

[Response Begins]

We calculated an interquartile range to determine if there were practically meaningful differences in performance measure scores between managed care plans. The IQR is calculated as the rate for the 75<sup>th</sup> percentile minus the rate for the 25<sup>th</sup> percentile. To determine if the difference between scores were statistically significant, we conducted an independent sample t-test between two randomly selected plans below the 25<sup>th</sup> percentile and above the 75<sup>th</sup> percentile. If the p value of the t-test is less than 0.05, then the performance measure scores of the plans are significantly different.

[Response Ends]

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

*Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.*

[Response Begins]

**Table 11. Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization, Variation in Performance Across NYS Medicaid Health Plans, 2018 NYS Medicaid Discharges**

	N	Min	P10	P25	Mean	Median	P75	P90	Max	IQR	P value
Engagement in Community-Based Mental Health Care After a Mental Health Hospitalization Rate	32	11%	25%	41%	46%	49%	53%	58%	66%	12%	p < 0.001

N = Number of plans reporting

IQR = Interquartile range

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

**[Response Ends]**

**2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.**

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

**[Response Begins]**

There was a difference of 55% between the lowest and highest scoring plans. The interquartile range, or difference between the 75<sup>th</sup> and 25<sup>th</sup> percentiles, was 12%. The p value from the t test was p<0.0001, which means that the difference between the two randomly selected plan scores was statistically significant.

**[Response Ends]**

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

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

**[Response Begins]**

This measure is based on NYS administrative Medicaid data. Claims are used to identify discharges and community-based follow-up visits. It is not expected that a large proportion of claims would be missing because they must be submitted to NYS in order for managed care plans and services providers to receive payment. There is a data lag of about six months in NYS claims. Our analysis used claims data from calendar year 2018 that was extracted after the six-month claims lag period had passed. Performance measures in NYS are always calculated at least six months after the end of the measurement period to minimize the effect of claims lag. When performance measures are calculated, the numerator and denominator counts for the current year are compared to previous years to check for any aberrations that could indicate missing data.

**[Response Ends]**

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

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

**[Response Begins]**

It is not possible to know how many claims were not submitted in Calendar Year 2018 because we have no secondary source of the number of services provided. The NYS Medicaid Data Warehouse and NYS Office of Mental Health monitor the number of claims submitted monthly. If there are aberrations in the number of claims compared to prior months, the differences are investigated and changes are made if it is determined that data is missing or was duplicated.

**[Response Ends]**

**2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.**

*In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.*

**[Response Begins]**

If aberrations in the monthly number of claims submitted are found, NYS will identify and correct the source of the error. When performance measures are calculated, if quality checks find differences of more than 5-10% in plan level denominator sizes, NYS OMH will identify the source of the difference and correct any data errors found before results are reported.

**[Response Ends]**

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

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

**[Response Begins]**

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

**[Response Ends]**

**2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.**

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

**[Response Begins]**

**[Response Ends]**

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

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

**[Response Begins]**

**[Response Ends]**

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

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

**[Response Begins]**

**[Response Ends]**

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

**[Response Begins]**

Yes, the measure uses exclusions.

**[Response Ends]**

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

*Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?*

**[Response Begins]**

This measure excludes:

1. Discharges who were dually enrolled in Medicaid and Medicare.
2. Discharges in hospice or using hospice services anytime during the measurement year.

Discharges who were dually enrolled in Medicaid and Medicare were excluded because this is a measure for Medicaid managed care plans and the entirety of services that Medicare/Medicaid duals receive cannot be reliably derived from Medicaid claims data. We would miss discharges or follow-up visits that are reported in Medicare if dual enrollees were include in the measure.

Discharges who received hospice services during the measurement year were excluded because the focus of care and criteria for quality care for this population may be different than the population not receiving hospice services. Engagement in community-based mental health care may not be a priority for individuals receiving hospice.

**[Response Ends]**

**2b.17. Provide the statistical results from testing exclusions.**

*Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.*

**[Response Begins]**

These exclusions are common for Medicaid performance measures, including the NCQA HEDIS measure Follow-Up After Hospitalization for Mental Illness, and were not tested.

**[Response Ends]**

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

*In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.*

**[Response Begins]**

Excluding Medicare and Medicaid duals is necessary because we cannot use Medicaid claims data to derive all of the services received during the measurement year. It enhances the fairness of the measure to exclude discharges that may have services missing in Medicaid data.

Excluding discharges who received hospice services during the measurement year is necessary because the focus of care and criteria for quality care for this population may be different than the population not receiving hospice services. It enhances the fairness of the measure to exclude discharges where engagement in community based mental health care may not be a priority.

**[Response Ends]**

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

**[Response Begins]**

No risk adjustment or risk stratification

**[Response Ends]**

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

**[Response Begins]**

**[Response Ends]**

**2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.**

**[Response Begins]**

**[Response Ends]**



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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

**2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter “N/A” for questions about the statistical risk model discrimination and calibration statistics.**

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

Not applicable, a statistical risk model was not used.

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

### 3. Feasibility

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

---

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

**3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.**

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

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

[Response Begins]

[Response Ends]



## 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.

---

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

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

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

*A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*

[Response Begins]

[Response Ends]

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

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

**4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

## 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.

---

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

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

**NOTE: If there are no related measures, please select N/A.**

*(Can search and select measures.)*

[Response Begins]

[Response Ends]

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

**NOTE: If there are no competing measures, please select N/A.**

*(Can search and select measures.)*

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

[Response Begins]

[Response Ends]

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

*Provide analyses when possible.*



[Response Begins]

[Response Ends]

## Appendix

Supplemental materials may be provided in an appendix.:

## Contact Information

**Measure Steward (Intellectual Property Owner):** New York State Office of Mental Health

**Measure Steward Point of Contact:** Ronsani, Adrienne, [adrienne.ronsani@omh.ny.gov](mailto:adrienne.ronsani@omh.ny.gov)

**Measure Developer if different from Measure Steward:** New York State Office of Mental Health

**Measure Developer Point(s) of Contact:** Ronsani, Adrienne, [adrienne.ronsani@omh.ny.gov](mailto:adrienne.ronsani@omh.ny.gov)

## Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]

[Response Ends]

2. List the workgroup/panel members' names and organizations.

*Describe the members' role in measure development.*

[Response Begins]

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]

[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]

[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]

[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]

[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".

[Response Begins]

[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate "N/A".

[Response Begins]

[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".

[Response Begins]

**[Response Ends]**