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

Purple text represents the responses from measure developers.
Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 3590

Corresponding Measures:

De.2. Measure Title: Continuity of Care After Receiving Hospital or Residential Substance Use Disorder (SUD) Treatment

Co.1.1. Measure Steward: RTI International

De.3. Brief Description of Measure: Percentage of Medicaid discharges, ages 18 to 64, being treated for a substance use disorder (SUD) from an inpatient or residential provider that received SUD follow-up treatment within 7 or 30 days after discharge. SUD follow-up treatment includes outpatient, intensive outpatient, or partial hospitalization visits; telehealth encounters; SUD medication fills or administrations; or residential treatment (after an inpatient discharge). Two rates are reported: continuity within 7 and 30 days after discharge.

1b.1. Developer Rationale: Remaining in addiction treatment for an adequate period is critical for recovery (National Institute on Drug Abuse, 2018). Patients often drop out of SUD treatment during transitions from inpatient and residential setting to outpatient settings (Harris, et al. 2006; Naeger et al., 2016; Ali et al., 2016; Reif, et al., 2017; Rubinsky et al., 2017; Liu et a., 2020). Studies with Medicaid beneficiaries document that, on average, only 25% of them receive a post-discharge follow-up within 14 days of a residential or inpatient SUD stays (although rates vary significantly among states and providers) (Harris et al., 2006; Naeger et al., 2016; Ali et al., 2016; Reif et al., 2017; Rubinsky et al., 2017; Liu et a., 2020).

Research finds that post-discharge follow-up after discharge from an inpatient or residential SUD stay is associated with better outcomes, such as reduced mortality (Harris et al., 2015; Paddock et al., 2017; Schmidt et al., 2017), readmissions (Mark et al.., 2013, Reif et al., 2017), substance use and improved retention in treatment (Demarce et al., 2018, Garner et al., 2010). The use of a performance measure to support post-discharge follow-up after discharge from an inpatient or residential SUD stay can support quality improvement efforts.

NQF endorsed the metric Continuity of care after inpatient or residential treatment for substance use disorder (NQF 3453) for use in Medicaid programs and health plans to stimulate improvement in post-discharge

continuity rates. The measure is defined as the percentage of discharges from inpatient or residential treatment for SUD for Medicaid beneficiaries, ages 18–64, followed by a SUD treatment service.

A provider level measure is needed in addition to the Medicaid and health plan level measure for several reasons. First, there is significant variation among providers in post-discharge follow-up after inpatient and residential SUD treatment (Stein et al., 2009; Rubinsky et al., 2018). Creating a provider-level measure allows states, payers, policymakers, and others to target quality improvement to providers that need it. Second, provider-level measures can reveal why some providers have lower follow-up rates than others and identify solutions. Interventions to improve post-discharge follow-up include inpatient addiction consults (Englander et al., 2019), scheduling outpatient appointments before discharge, starting patients on medications to treat opioid use disorder mediations before discharge, using peer navigators, and facilitating obtaining housing and other social supports (Bassuk et al., 2016, Wakeman et al. 2017, Manuel et al., 2017; Liebschutz et al., Wang et al., 2020), Third, NQF SUD measures are being used at the provider-level in Centers for Medicare and Medicaid (CMS) demonstrations, such as in behavioral health home demonstrations; however, they have not been endorsed at the provider level (CMS, 2019). Finally, some states and private health plans are already using the measure at the provider level, for example, New York's Office of Addiction Services and Support is using the measure to help programs improve follow-up rates, and the measures are being reported for New York, Massachusetts, and West Virginia as part of the Shatterproof Atlas portal.

This proposed provider-level measure - Continuity of care after inpatient or residential treatment for substance use disorder - has a similar logic model endorsed at the Medicaid program level. The logic model, and associated evidence, indicates that this measure could help to reduce hospital readmissions (Mark et al., 2013; Reif et al., 2017), decrease substance use and relapse (DeMarce et al., 2008; Garner et al., 2010), and lower mortality (Harris et al., 2015; Paddock et al., 2017; Schmidt et al., 2017). Potential benefits to society include reduced costs related to lower crime rates and decreased health care expenditures (Popovici, French, & McKay, 2008; Heslin et al., 2015).

S.4. Numerator Statement: Medicaid discharges, ages 18 to 64, with a principal/primary substance (SUD) diagnosis treated at an inpatient or residential provider that received SUD follow-up treatment within 7 or 30 days after discharge. SUD treatment includes outpatient, intensive outpatient, or partial hospitalization visits; telehealth encounters; or SUD medication fills or administrations; or residential treatment (after an inpatient discharge. Two rates are reported: continuity within 7 and 30 days after discharge.

S.6. Denominator Statement: The denominator are Medicaid beneficiaries, ages 18-64, discharged from inpatient or residential provider with a principal diagnosis of SUD on the inpatient/residential treatment encounter claim.

S.8. Denominator Exclusions: Dual eligible Medicare/Medicaid beneficiaries are excluded. Rationale: Individuals who are covered under Medicare would receive coverage for follow-up treatment medications (e.g. opioid use disorder medications) under Medicare Part D and Medicare Part D claims are not captured in Medicaid claims databases. Therefore, follow-up treatment would be missed.

De.1. Measure Type: Process

S.17. Data Source: Claims, Enrollment Data

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?

Preliminary Analysis: Maintenance of Endorsement New Measure

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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 is a new claims-based, facility level measure that assesses the percentage of Medicaid discharges, ages 18 to 64, being treated for a substance use disorder (SUD) from an inpatient or residential provider that received SUD follow-up treatment within 7 or 30 days after discharge. SUD follow-up treatment includes outpatient, intensive outpatient, or partial hospitalization visits; telehealth encounters; SUD medication fills or administrations; or residential treatment (after an inpatient discharge). Two rates are reported: continuity within 7 and 30 days after discharge.
- Developer provides a logic model depicting the relationship between structural and process changes to improve follow up after a SUD-related discharge and better outcomes for patients.
- Developer summarizes the existing follow up literature as follows:
    - Research suggests that follow up after discharge from inpatient or residential SUD stays is associated with better outcomes, such as reduced mortality, readmissions, substance use and treatment abandonment.
    - Developer cites 12 studies to this effect, mostly adjusted retrospective analyses with some RCTs.
    - Associations are directionally consistent, moderate quantity, moderate quality.

Questions for the Committee:

- The developer's evidence provided evidence supportive of the measure focus which was all directionally positive. Is the Committee aware of evidence that suggests that follow up after SUD inpatient care is not effective?

Guidance from the Evidence Algorithm

(Box 1) Process measure (Box 3) Systematic reviews but not graded (Box 7/8) All studies summarized (Box 9) Benefits outweigh negative effects MODERATE

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

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

Maintenance measures – increased emphasis on gap and variation

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

- Developer provides performance scores based on Medicaid claims data from 623 providers who treated 81,720 beneficiaries demonstrate a significant performance gap.
- The median 7-day follow-up rate was 11%, and the median 30-day follow-up rate was 24%.
- There was also significant variation among providers.
    - Observed scores for the 7-day follow-up ranged from 0% to 99%, with a mean of 19% and SD of 22%.
    - Observed scores for the 30-day follow-up ranged from 0% to 99%, with a mean of 29% and SD of 13%.
- Summary Data of Observed Scores:

Measure	n	Mean	SD	Min	10th	25th	50th	75th	90th	Max
7 Days	623	19%	22%	0%	0%	4%	11%	25%	52%	99%
30 Days	623	29%	13%	0%	3%	10%	24%	43%	63%	99%

Disparities

- Developer notes disparities by gender and race in continuity of care after SUD treatment in a hospital or residential setting.
    - Developer analysis suggests males were less likely to receive follow-up care (15% versus 23%).
    - Blacks were less likely than Whites to receive follow-up care (22% versus 9%).

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:
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.

- Solid evidence of continuity of care in SUD at provider level of analysis.
- Evidence is sufficient.
- there's evidence to support significance
- This evidence in this measure directly relates to quality and outcomes. The 25% rate of follow-up after discharge is alarmingly low.
- Evidence is moderate
- The evidence relates directly to the measure. The developer includes evidence regarding current utilization of treatment after inpatient or SUD residential treatment. The process relates to whether these patents are seen for follow-up treatment post discharge, and information from this measure could be used to increase provider practices that would facilitate treatment post discharge and improve overall care. No systematic review provided. This is a new measure. Measure derived from claims and enrollment data, not from patient report. I am not aware of any evidence that suggests follow-up after SUD inpatient care is not effective.
- Evidence applies directly to measure.
- Similar to other follow up measures, why the intervals chosen? Is 7 days better than 30? Also, is this at the facility or the provider level--I am confused from the description and materials. If at an individual level, I again don't see the evidence. Isn't this a system's issue that should be accountable at a system's level? Any case, show me the data--what is presented is true, true and only distantly related.
- Evidence applies directly to the process measure. The evidence supports the benefits continuity of SUD care. The process measure has the potential to improve patient outcomes such as reduced mortality, readmissions, and substance use. I am not aware of any new information that changes the evidence base for this measure that has not been cited in the submission.
- Evidence seems sufficient. Not aware of evidence suggesting follow-up after SUD is not effective.
- Empirical data cited by submission applies directly to the measure and desired outcomes (e.g. follow-up care after inpatient hospitalization for SUD). SUD is a growing problem in the United States and literature cites evidence to support follow-up outpatient care to reduce readmission rate.
- Process
- NA
- The evidence, while I agree moderate in quantity and quality, do support this measure.

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?

- demonstrated room for improvement.
- Sufficiently demonstrates performance gaps.

- overall low rates of adherence with SD's that suggest variation
- The submission notes significant gaps in care based on gender and race. Data on disparities was provided.
- Performance Gap is High
- Yes, performance data was provided for individual providers. Median rates for 7-day follow-up and 30-day follow-up were low, and disparities exist in gender and race for continuity of care. Men were less likely than women to have continuity of care, and Blacks were less likely than Whites.
- Significant performance gap present.
- There are gaps.
- There is high variability across providers warranting a national performance measure that can help improve these gaps in care. Disparities were noted, specifically men are less likely to receive follow-up care than women. Blacks are less likely to receive follow-up care than Whites.
- Does seem to be sizable variability across providers, suggesting major room for improvement. In terms of disparities, males were less likely to receive follow-up care than females and blacks were less likely to receive follow-up care than whites suggests it IS important to address disparities as performance focus.
- There is a high gap in care, as the mean 7-day follow-up rate was 19% and the 30-day rate was 29%. Data on gender and race were provided by the measure developer and indicate gaps in care that could be further studied.
- yes there are opportunities of improvement with men and blacks.
- Plenty of opportunity for improvement shown. Data shows that blacks less likely to receive follow up than whites and males less likely than females.
- There is a notable performance gap and noted disparities that this measure would help address.

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 Measure Evaluation

NQF Staff Evaluation Summary:

Reliability

- Developer conducted several score-level reliability analyses, including:
    - Parametric ANOVA with effect size calculations
        - 7-day follow up: $F = 63.9$; $\eta 2 = 0.25$; $\omega 2 = 0.24$
        - Facilities: $F = 56.6$; $\eta 2 = 0.22$; $\omega 2 = 0.22$
    - Intra-unit reliability (IUR) – 0.94 for 7-day follow up; 0.93 for 30-day follow up
    - Beta-binomial signal to noise analysis:

Measure	n	Mean	SD	Min	10th	25th	50th	75th	90th	Max
7-day	623	0.94	0.07	0.67	0.83	0.92	0.98	1.00	1.00	1
30-day	623	0.93	0.07	0.71	0.82	0.90	0.96	0.99	1.00	1

- Staff concurs with developer assessment of the results: The reliability testing results suggest that the measure is highly reliable. The F-statistic for the signal-to-noise ratio indicates that the measure scores are significantly different while subsequent reliability statistics (the IUR and Adam's rho) indicate a large effect size, empirically substantiating that the measure can discern underlying performance between providers.

Validity

- Convergent validity is established by empirically showing that measures that are conceptually related are statistically correlated to one another.
- Pearson product moment correlation coefficients were calculated between the measure and a measure of medication for opioid use disorder (MOUD) continuity of therapy

- The correlation between the SUD Follow-up measure at 7-day follow-up and the MOUD measure was r=0.39 (p<0.001)
- The correlation between the SUD Follow-up measure at 30-day follow-up and the MOUD measure was r=0.39 (p<0.001).

- Convergent validity testing was conducted using a common method for conceptually similar measures, producing results that were statistically significant, directionally appropriate and of moderate strength.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The NQF staff or is satisfied with the reliability testing for the measure. Does the Committee agree with the staff assessment of the reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The NQF staff or is satisfied with the validity testing for the measure. Does the Committee agree with the staff assessment of the validity?

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

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?

- above average
- No concerns.
- The specifications could make it problematic to identify where to target intervention to improve. For example, inpatient or residential provider are lumped and clinical severity and complexity could vary, also follow-up from hospital d/c includes going to residential? SUD follow-up care is also very broadly defined--this might be OK, but includes potentially medication only?
- I concur with the NQF staff assessment that reliability testing is high. I have no concerns about consistent implementation of this measure.
- Does the numerator include individuals treated with Narcan for overdose in the ED ?
- Data elements are clearly defined. Codes with descriptors are provided and the steps are clear. I do not have concerns that this measure can be consistently implemented as reliability estimates are high.
- Numerator includes several different types of visits, that may be confusing and result in inconsistent implementation.
- OK

- Data elements are clearly defined, with adequate sampling, and good reliability demonstrated through a standard methodology. I do not have concerns about the likelihood that this measure can be consistently implemented.
- Measure specifications seem adequate.
- I have no concerns about implementation as the data used to measure are claims data.
- No concerns
- The decision to exclude patients with Medicare essentially makes the population younger which has the potential to skew any correlations to mortality and possibly make the results better than they would otherwise be.
- I agree with the reliability testing done and have no concerns.

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

- no
- No.
- similar approach by RTI, no major concerns
- No, I concur with the NQF staff assessment.
- Highly reliable
- I do not.
- See above.
- ok
- No because ANOVA, signal-to-noise, IUR, Adam's rho, demonstrated underlying performance between providers were all were within acceptable range and indicate large effect size.
- Score-level reliability analyses seem adequate.
- No
- No concerns.
- reliability seems OK
- No

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

- only evaluated vs medication continuation. would have been useful to compare validity vs ASI or non-medication outcome measure.
- No.
- convergent validity r=.39. Similar to other developers, no rigorous testing that adherence is associated with improved clinical outcomes
- No
- Moderate
- I do not.
- No concerns.
- pearson r of 0.39 seems rather low, moderate a stretch.

- I do not.
- Empirical validity tests seem adequate. No concerns.
- No
- No concerns.
- Validity seems adequate
- No concerns.

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?

- It's unclear if transfer to mental health, medical or hospital is excluded.
- Acceptable.
- given data source, unable to stratify by most social determinants of health
- I am concerned about the exclusion of dual eligible beneficiaries. Medicare Part B largely excludes addiction treatment, while Medicaid in many states covers residential care. So while Part D covers MAT, most other addiction treatment is covered by Medicaid for dual eligibles age.
- no concerns
- Exclusions are solely based on payer type and have been established to reduce the likelihood of missing data.
- No concerns.
- OK
- No concerns about the patients excluded from the measure. Risk adjustment - n/a.
- No concerns. Risk adjustment or stratification N/A.
- N/A
- No risk adjustment in this measure.
- Excluding Medicare population will probably artificially lower mortality a bit. Not a fatal flaw but one that may make comparison to other literature fraught with peril.
- None.

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?

- none

- I am satisfied with the testing of the measure's validity.
- no
- I concur with the NQF finding of no threat to validity.
- No concerns
- This measure includes data on follow-up visits post discharge from inpatient or SUD treatment facilities. Analyses of this measure indicate that there are meaningful differences among providers and that the care received differs among population subgroups. Yes, measure is specified precisely indicating comparable results. The developer reports rare instances of missing data, such that it should not have an impact on the measure.
- No concerns.
- I wonder about the ability to gather comprehensive data given carve outs and potential barriers to accurate reporting
- The F-statistic indicates that the measure scores are significantly different, IUR and Adam's rho indicates a statistically significant difference in performance between providers, and Eta-squared and omega-squared values indicate effect size. Comparability - n/a. Missing data - reported to be rare and does not impact the measure.
- No concerns. Missing data not expected to be an issue. Analysis addresses meaningful differences. Comparability N/A
- Measure assesses meaningful differences, but not from multiple data sets. Missing data does not constitute a threat to validity.
- No threat to validity.
- I wonder if it might be helpful for us to compare these results to the results of the similar Medicaid and health plan level measure. Systems of care generally do not the resources to devote to ensuring their patients have care navigators/care manager to aid patients when transitioning to a new provider at a different level of care. Nor do they have resources devoted to analyzing their results and improving them. Those systems of care that have their own follow up levels of care/resources are likely to have an advantage of getting their patients to actually follow up (indicating truly improved follow-up) AND are more likely to have the info to more reliably bill for their follow up care (helping their finances but not necessarily indicating better quality).
- A few of the ICD codes used to identify SUD may have issues, especially opioid dependence as this may capture people with physiological dependence but not use disorder.

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.
    - Measure uses claims, where data elements are routinely generated and used during care delivery. Coded by someone other than person obtaining original information. All data elements are in defined fields in a combination of electronic sources.

Questions for the Committee:

- Does the Committee have any feasibility concerns for the measure?

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?

- appears to use existing claims data--but not stated how that is collected.
- Feasible.
- feasible using claims data
- I concur with the NQF staff finding of high feasibility.
- State owned psychiatric MH inpatient stays are not in the Medicaid claims data
- Data are generated during usual care and coded after delivery such as diagnosis. All data are available in electronic form. I do not have concerns about how the data can be put into use.
- No concerns.
- Feasible
- No concerns regarding feasibility. Data elements are generated or collected by healthcare personnel during the provision of care and all data elements are in defined fields in electronic claims.
- No concerns about feasibility – data elements are all part of regular claims.
- No concerns about feasibility since the measure uses claims data that are already routinely collected
- No concerns .
- No concerns
- No concerns.

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

- New York Office of Addiction Supports and Services, Shatterproof ATLAS
- Geographic area and number and percentage of accountable entities and patients included: New York state (approximately 90 addiction treatment facilities), Shatterproof ATLAS (approximately 130 addiction treatment providers across 3 states, New York, Massachusetts, and West Virginia).
- Level of measurement and setting: Specialty addiction treatment facility

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

- Developer notes that "the measure was developed with feedback from state Medicaid programs, commercial health plans, addiction treatment providers, patients, families, and other experts. Experts reviewed the measure as part of an NQF sponsored Strategy session. Focus groups were held with providers, patients, and families to obtain feedback on the measures. One Medicaid program and one commercial health plan helped to test and refine the initial specification. The measure was then implemented by three Medicaid programs as part of Shatterproof's Addiction Treatment Locator, Assessment, and Standards (ATLAS) Platform. New York State's Office of Addiction Supports and Services has integrated the measure into its quality improvement activities."

Additional Feedback: N/A

***Questions for the Committee**:*

- How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

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

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

4b. Usability 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

- Measure has not been implemented

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

Unexpected findings (positive or negative) during implementation: None identified.

Potential harms: None identified.

Additional Feedback: N/A

***Questions for the Committee**:*

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- Do the benefits of the measure outweigh any potential unintended consequences?

Preliminary rating for Usability and use: ☐ **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?

- Useable as both quality and performance indicator.
- Yes
- being used, but not clear if data were provided to support good data quality
- The submission notes that there is currently no public reporting. The measure appears to be widely used by Medicaid agencies for accountability.
- yes.
- The measure is not being publicly reported and is being considered for initial endorsement. Data will be presented in a portal for those who have access. There is a credible plan for quality improvement through providing results back to providers. It is currently being used for external benchmarking for facilities and internal to specific organizations. It aims to improve treatment post discharge for SUD. There is a credible plan for implementation and the measure is currently in use for quality improvement. Yes, users are being presented data in a portal and are offered technical assistance material and training. Feedback has been received from stakeholders "Medicaid programs, commercial health plans, addiction treatment providers, patients, families and other experts" as well as experts from NQF and feedback from focus groups. Feedback is considered when incorporating changes.

- New measure.
- I still don't understand what has been learned through the NY experience with Shatterproof ATLAS
- Measure is not publicly reported but is currently being used and will be more broadly used in accountability programs. Feedback was garnered in the development phase of this measure from state Medicaid programs, commercial health plans,~130 addiction treatment providers, patients, families, and other experts.
- No concerns. New York State's extensive vetting with providers, patients, and other stakeholders lends support for its use. Having a measure at the provider/clinician level is important to continuous improvement.
- N/A, as this measure is not in use; however developer states that measure was developed using feedback from state Medicaid plans and others using similar measures and during an NQF session to gather feedback.
- Yes
- The MA/health plan level measure is already being used and this one is likely to follow suit.
- No concerns. It appears feedback was well-received for the development of this measure.

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.

- appears useable as demonstrated by New York use of similar measure.
- I have some concerns about who "owns" the process when patient transitions from one provider entity to another in transitioning care from residential to outpatient. Additionally, there may be risk of cherry picking as well as inappropriate use of incentive or coercion for OP participation.
- see concerns about specifications and data interpretation
- Given the dismal rate of follow-up after discharge from residential care, this measure is an important tool in improving quality and outcomes.
- benefits outweigh the harms
- If treatment post discharge from inpatient and SUD treatment centers improves outcomes, this measure can be used to assist providers in understanding their follow-up rates. Yes, there is a clear rationale for how performance results could be used to provide high quality, efficient healthcare. Since the measure relies on claims data for services already rendered and will provide feedback to providers that could support quality improvement, there are no unintended consequences of reviewing the data. Instead, there is potential to improve care, resulting in better outcomes for the individual and societal impacts such as reduced costs. One comment related to risk is the quality of the follow-up treatment services. How do providers and facilities determine quality to ensure they're sending patients to qualified providers and sites?
- Providers might treat only patients perceived to be compliant with f/u.
- Benefits outweigh harms
- The performance results can be used to further the goal of high-quality, efficient healthcare by monitoring the rate of continuing care, identifying gaps in care transitions, and encouraging states to

put improvements in place. The potential benefits of continuity of care include decreased rates of substance use and relapse, fewer readmissions for inpatient treatment, lower risk of death, less involvement in criminal justice, and improved employment outcomes. I believe the potential benefits (namely decreased mortality) outweigh the potential harms in this measure.

- No concerns.
- None identified; follow-up care for SUD is an important issue and benefits of a measure for this outweigh potential harms.
- None
- Rationale is credible
- Benefits outweigh any theoretical harms.

Criterion 5: Related and Competing Measures

Related or competing measures

Developer notes the following related measure:

- 3453 : Continuity of Care after Inpatient or Residential Treatment for Substance Use Disorder (SUD)

Harmonization

- The measure is harmonized with the parallel measure that was developed for use at the health plan or Medicaid program level (NQF 3453).
    - In both measures, the population is Medicaid beneficiaries age 18 – 64.
    - The same diagnosis codes are used to identify substance use disorders. The same services and procedures are included to define follow-up treatment.
- While NQF# 3453 examines post-discharge follow-up at 7 and 14 days, developer is proposing that #3590 Continuity of Care After Receiving Hospital or Residential Substance Use Disorder (SUD) Treatment be reported at 7 and 30 days.
    - This is because review of the evidence and discussion with addiction professionals supported measuring follow up for an extended period of time.
    - Also, NQF# 3488 Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence, NQF#3489 Follow-up after Emergency Department Visit for MH, and NQF# 0577 Follow-up after a hospitalization for a mental illness indicated that these measures are reported at 7 and 30 days.
    - Finally, HEDIS has implemented NQF #3453 as measuring following-up at 7 and 30 days, not at 7 and 14 days (https://www.ncqa.org/wp-content/uploads/2019/02/20190208_06_FUI.pdf). This difference between using a 14- or 30-day follow-up does not impact interpretability or data collection burden.

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?

- yes. appears addressed in staff notes
- harmonized with NQF 3453
- The submission notes that the measure is harmonized with the parallel measure that was developed for use at the health plan or Medicaid program level (NQF 3453).
- harmonized
- One other measure is listed by the developer and an explanation for how the two have been harmonized is included.
- no.
- Why don't we just use the current approved measure and then consider whether to approve at the clinician or facility level?
- 3453 - harmonized, 3488, 0577
- While there are related measures, they do not appear to be competing in any way, and measure specifications (for 3453) are adequately harmonized. Supportive of proposal to report this new measure at 7 and 30 days.
- Measure has been harmonized with NQF3453, using same coding; however current measure reports results at 7 and 30 days follow-up rather than 7 and 14 days follow-up to be consistent with clinical practices and other related measures (e.g. follow-up after discharge for mental illness).
- None
- The MA/health plan level related measure might give us pause for thought: what can we learn about resources and leverage that could aid the delivery system as well as the plans?
- This measure is well-harmonized with existing ones and appropriate given feedback from addiction professionals to extend the period of time.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/15/2021

- No NQF Members have submitted support/non-support choices as of this date.
- No Public or NQF Member comments submitted as of this date.

NQF Staff Scientific Acceptability Evaluation

Measure Number: NQF 3590

Measure Title: Continuity of Care After Receiving Hospital or Residential Substance Use Disorder (SUD) Treatment

Type of measure:

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

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

Data Source:

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

Level of Analysis:

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

   Submission document: Items S.1-S.22

2. **Briefly summarize any concerns about the measure specifications.**
   - None identified

RELIABILITY: TESTING

Submission document: Specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**

   ☒ **Yes** ☐ **No**

5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☐ **Yes** ☐ **No**

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

   Submission document: Testing attachment, section 2a2.2

   - Developer conducted several score-level reliability analyses, including:
     - Parametric ANOVA with effect size calculations
       - 7-day follow up: $F = 63.9$; $\eta 2 = 0.25$; $\omega 2 = 0.24$
       - Facilities: $F = 56.6$; $\eta 2 = 0.22$; $\omega 2 = 0.22$
     - Intra-unit reliability (IUR) – 0.94 for 7-day follow up; 0.93 for 30-day follow up
     - Beta-binomial signal to noise analysis:

Measure	n	Mean	SD	Min	10th	25th	50th	75th	90th	Max
7-day	623	0.94	0.07	0.67	0.83	0.92	0.98	1.00	1.00	1
30-day	623	0.93	0.07	0.71	0.82	0.90	0.96	0.99	1.00	1

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

   Submission document: Testing attachment, section 2a2.3

   Staff concurs with developer assessment of the results: The reliability testing results suggest that the measure is highly reliable. The F-statistic for the signal-to-noise ratio indicates that the measure scores are significantly different while subsequent reliability statistics (the IUR and Adam's rho) indicate a large effect size, empirically substantiating that the measure can discern underlying performance between providers.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

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

10. **OVERALL RATING OF RELIABILITY** (considering 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.**

    (Box 1) → Measure specifications precise, unambiguous, and complete (Box 2) → Empirical testing conducted using statistical tests → (Box 4): Reliability testing conducted with computed performance measure scores → (Box 5): Method described and appropriate for assessing the proportion of variability due to real differences among measured entities → (Box 6a) HIGH

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

- None identified by staff

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

    Submission document: Testing attachment, section 2b4.

    - None identified by staff

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.

    - None identified by staff

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

    Submission document: Testing attachment, section 2b6.

    - None identified by staff

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

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

    - Convergent validity is established by empirically showing that measures that are conceptually related are statistically correlated to one another.
    - Pearson product moment correlation coefficients were calculated between the measure and a measure of medication for opioid use disorder (MOUD) continuity of therapy
        - The correlation between the SUD Follow-up measure at 7-day follow-up and the MOUD measure was r=0.39 (p<0.001)
        - The correlation between the SUD Follow-up measure at 30-day follow-up and the MOUD measure was r=0.39 (p<0.001).

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

    Submission document: Testing attachment, section 2b2.3

    - Convergent validity testing was conducted using a common method for conceptually similar measures, producing results that were statistically significant, directionally appropriate and of moderate strength.

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 considering 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.**

    (Box 1)-All potential threats to validity assessed → (Box 2) Empirical validity testing conducted using the measure as specified and appropriate statistical testing → (Box 6) Validity testing conducted with computed performance measure scores of each measured entity → (Box 7) Method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships → (Box 8b) Moderate certainty or confidence that the performance measure scores are a valid indicator of quality- MODERATE

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.**

Developer Submission

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

SUD-Follow-up_Evidence_11_19_2020.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*)**:**

Measure Title: Continuity of Care After Receiving Hospital or Residential Substance Use Disorder (SUD) Treatment

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

Date of Submission: 10/30/2020

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):

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*):

☐ Process: Measures receipt of SUD treatment within 7 days and/or 30 days after inpatient discharge or discharge from a hospital or residential treatment program for treatment of a SUD.

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

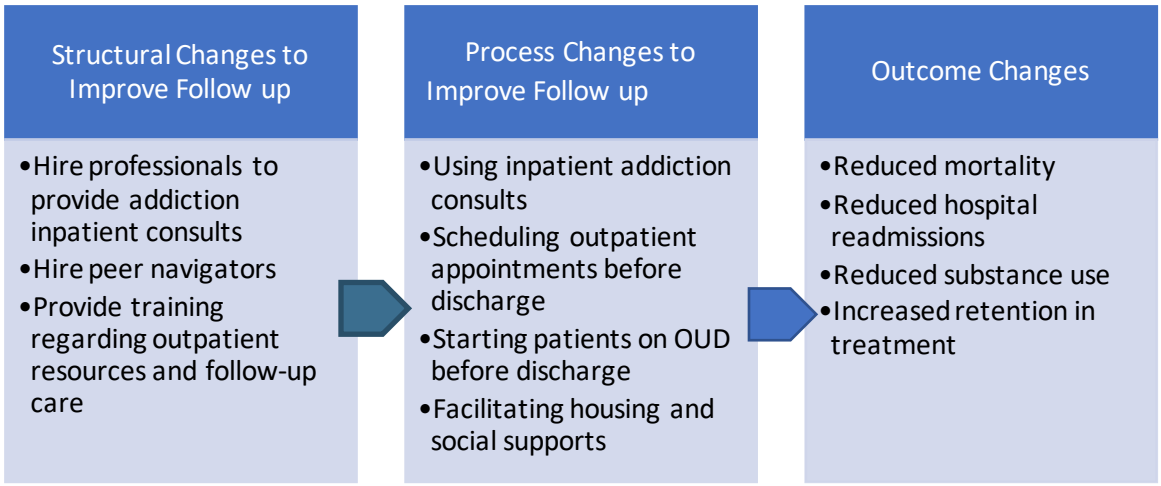
☐ Structure:

☐ Composite:

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.

Figure 1. Logic model explaining how the measures will lead to improved follow up care and better outcomes



Remaining in addiction treatment for an adequate period is critical for recovery (National Institute on Drug Abuse, 2018). Patients often drop out of SUD treatment during transitions from inpatient and residential setting to outpatient settings (Harris, et al 2006; Naeger et al., 2016; Ali et al., 2016; Reif, et al. , 2017; Rubinsky et al., 2017; Liu et a., 2020). Studies with Medicaid beneficiaries document that, on average, only 25% of them receive post-discharge follow-up within 14 days of a residential or inpatient SUD stays, with significant differences among states and providers) (Harris et al, 2006; Naeger et al., 2016; Ali et al., 2016; Reif et al, 2017; Rubinsky et al., 2017; Liu et a., 2020).

Research finds that post discharge follow-up after discharge from an inpatient or residential SUD stay is associated with better outcomes, such as reduced mortality (Harris et al., 2015; Paddock et al., 2017; Schmidt et al, 2017), readmissions (Mark et al., 2013, Reif et al., 2017), substance use and improved retention in treatment (Demarce et al., 2018, Garner et al., 2010). Performance measures can help stimulate improvement in post-discharge follow-up rates.

NQF endorsed the metric *Continuity of care after inpatient or residential treatment for substance use disorder (*NQF 3453) for use in Medicaid programs and health plans to stimulate improvement in post-discharge continuity rates. The measure is defined as the percentage of discharges from inpatient or residential treatment for SUD for Medicaid beneficiaries, ages 18–64, which were followed by a treatment service for SUD.

A provider level measure is needed in addition to the Medicaid and health plan level measure for several reasons. First, there is significant variation among providers' in post-discharge follow-up after inpatient and residential SUD treatment *(*Stein et al., 2009*,* Rubinsky et al*.,* 2018*).* Creating a provider-level measure allows states, payers, policymakers, and others to target quality improvement to providers that need it. Second, provider level measures can help to reveal why some providers to have lower follow-up rates than others and to identify solutions. Interventions to improve post-discharge follow-up include inpatient addiction consults (Englander et al., 2019), scheduling outpatient appointments before discharge, starting patients on medications to treat opioid use disorder mediations before discharge, using peer navigators,

and facilitating obtaining housing and other social supports (Bassuk et al., 2016, Wakeman et al 2017, Manuel et al., 2017; Liebschutz et al., Wang et al., 2020), Third, NQF SUD measures are being used at the provider-level in Centers for Medicare and Medicaid (CMS) demonstrations, such as in behavioral health home demonstrations, however, they have not been endorsed at the provider level (CMS, 2019). Finally, some states and private health plans are already using the measure at the provider level, for example, New York's Office of Addiction Services and Support is using the measure to help programs improve follow-up rates and the measures is being reported for New York, Massachusetts and West Virginia as part of the Shatterproof Atlas portal.

This proposed provider level measure - *Continuity of care after inpatient or residential treatment for substance use disorder* - has a similar logic model to that which was endorsed at the Medicaid program level. The logic model, and associated evidence, indicates that this measure could help to reduce hospital readmissions (Mark et al., 2013; Reif et al., 2017), decrease substance use and relapse (DeMarce et al., 2008; Garner et al., 2010), and lower mortality (Harris et al., 2015; Paddock et al., 2017; Schmidt et al, 2017). Potential benefits to society include reduced costs related to lower rates of crime and decreased health care expenditures (Popovici, French, & McKay, 2008; Heslin et al., 2015).

As the measure relies on administrative data, it should have low cost of adoption. However, there will still be costs for programmers to implement the measure. The measure may also result in some increase in healthcare costs. Costs may increase if inpatient and residential providers invest more resources in improving transitions of care and if more people receive treatment after discharge. These additional costs may be offset by reductions in SUD-related hospitalizations and emergency department visits.

A potential unintended consequence of this measure is that providers try to improve on the measure by selecting picking patients who are most likely to have post-discharge follow-up or provider may reduce admissions all together. These unintended consequences may be more likely if the measure is paired with a large financial incentive.

Measure Information

Measure description: Percentage of discharges from an inpatient or residential treatment for substance use disorder (SUD) for Medicaid beneficiaries, ages 18 to 64, which was followed by a treatment service for SUD. SUD treatment includes having an outpatient visit, intensive outpatient encounter or partial hospitalization, telehealth encounter, or filling a prescription or being administered or ordered a medication for SUD. (After an inpatient discharge only, residential treatment also counts as continuity of care.) Two rates are reported, continuity within 7 and 30 days after discharge.

Numerator: Discharges in the denominator with an outpatient visit, intensive outpatient encounter or partial hospitalization, telehealth encounter or filled a prescription for or were administered or ordered a medication for SUD.

Denominator: Adult Medicaid beneficiary discharges from inpatient or residential treatment for SUD with a principal diagnosis of SUD during from January 1 to December 15 of the measurement year.

Exclusions:

- The initial discharge and the admission/direct transfer discharge if the admission/direct transfer discharge occurs after December 15 of the measurement year.

Measure Implementation

Quality improvement

Supports quality monitoring and improvement at the state, program, or provider level

Benefits

Health care:

- Continuity of care after inpatient or residential treatment is generally low and varies significantly among providers within Medicaid programs; thus, there is much room for improvement in continuity of care

Health outcomes after inpatient or residential discharge:

Impact on clients

- Fewer hospital readmissions
- Less substance use and relapse
- Improved employment status
- Reduction in criminal justice activity
- Lower mortality

Impact on society

- Reduction in costs related to lower crime rates and lower health care costs

Health care/Medicaid savings:

- Lower costs as a result of continuity of care that helps to sustain a patient's gains from the initial treatment and to prevent relapses

Influencing Factors

- Health system issues
  - System organization and capacity
  - Location of necessary services

Costs and Unintended Consequences

Measure Implementation costs:

- Low cost to adopt measure because it relies on administrative data
- Cost for programmers to implement the measure

Cost of Improved Continuity:

- Increased resources spent by inpatient and residential SUD providers to improve post discharge follow-up
- Increased cost to Medicaid if more people receive SUD services after discharge
- These additional costs may be off-set by reduced readmissions for SUD and SUD-related

Unintended consequences:

- Providers may avoid clients whom they consider less likely to have continuity or avoid admitting clients with SUD.

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

N/A

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

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

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Not Applicable

Systematic Review	Evidence
Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	*
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.	*
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	*
Provide all other grades and definitions from the evidence grading system	*
Grade assigned to the **recommendation** with definition of the grade	*
Provide all other grades and definitions from the recommendation grading system	*
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	*
Estimates of benefit and consistency across studies	*
What harms were identified?	*
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	*

*cell intentionally left blank

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.

The evidence on which this measure is based is the result of a review of the literature of the effect of continuity of care/post-discharge follow-up after inpatient or residential treatment for substance use disorders.

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

The National Institute on Drug Abuse's *Principles of Drug Addiction Treatment* explains that "Remaining in treatment for an adequate period of time is critical for recovery." Moreover, they note that "research indicates that most addicted individuals need at least 3 months in treatment to significantly reduce or stop their drug use and that the best outcomes occur with longer durations of treatment." (National Institute on Drug Abuse, 2018). Patients often drop out of SUD treatment during transitions from inpatient and residential setting to outpatient settings (Harris, et al 2006; Naeger et al., 2016; Ali et al., 2016; Reif, et al. , 2017; Rubinsky et al., 2017; Liu et a., 2020).

Research finds that post discharge follow-up after discharge from an inpatient or residential SUD stay is associated with better outcomes, such as reduced mortality (Harris et al., 2015; Paddock et al., 2017; Schmidt et al, 2017), readmissions (Mark et al.., 2013, Reif et al., 2017), substance use and improved retention in treatment (Demarce et al., 2018, Garner et al., 2010) (See Table 1).

Mortality. Two studies found lower rates of death within one to two years following treatment for patients that had continuity of care services following discharge from residential SUD treatment, treatment engagement and detoxification (Harris et al., 2015; Paddock et al., 2017; Schmidt et al, 2017). In one study, patients who had continuity of care within 14 days of discharge from residential treatment had lower two-year mortality ratio (OR = 0.77, p<.008) (Harris et al., 2015). Another study found that patients with two or more follow-up treatment encounters within 30 days of the start of a treatment episode had lower rates of mortality in the one- and two-year periods following treatment (OR = 0.65, CI: 0.58, 0.74 & OR = 0.78, CI: 0.71, 0.85) (Paddock et al., 2017). A third study found a nearly 50% decrease in mortality in patients who followed up within 7 days of detoxification treatment or the end of detoxification-related prescriptions compared to those who did not follow-up within that time frame (OR = 0.65, P<0.001) (Schmidt et al., 2017).

Readmission. Studies find that continued care led to lower rates of readmissions after initial SUD treatment (Reif et al., 2017; Mark et al., 2013). One study found that, among a sample of Medicaid beneficiaries with an inpatient SUD treatment or residential detoxification hospital admission, those who received residential treatment after discharge had lower rates of readmission compared to beneficiaries with no-follow up (HR=0.05, p<.001) (Reif et al., 2017). Medication-assisted treatment for SUD after an inpatient or detoxification stay was similarly associated with lower rates of readmission (HR=0.61, p<.001) (Reif et al., 2017). Another study of Medicaid beneficiaries who received inpatient mental health or SUD treatment found that patients who received continued care treatment at community mental health centers after discharge had lower rates of readmissions (Mark et al., 2013).

Retention in Treatment and Reduced Substance Use. One study evaluated the efficacy of a continuing care adherence intervention consisting of contracting, prompting and reinforcing attendance in continued care (Demarce et al., 2008). The randomized study found that patients who received the adherence intervention after residential treatment had higher rates of 3-month follow-up as well as higher 1-year abstinence rates than patients who received standard treatment without continuing care adherence interventions (Demarce et al., 2008). Another study of adolescents treated in publicly funded long-term residential care facilities found similar results. Individuals who were randomized into a group to receive services to support continuing care after residential treatment were more likely to be abstinent at follow-up than patients who received standard care (Garner et al., 2010).

Table 1. Studies of Continuity/follow-up Impact on Mortality, Readmission, Retention in Treatment and Substance Use

Study	Design	Time After Discharge to Continuity of Care	Outcomes	Results
Harris et al. 2015	Veterans' Health Administration administrative data; cross sectional analysis (FY 2009), using propensity score weighted mixed effects regression models	Within 14-days after residential discharge	2-year mortality	76% of 10,064 patients had a continuity of care service within 14 days of discharge. Patients who had continuity of care had a lower 2-year mortality rate (OR= 0.77, p=0.008).
Reif et al. 2017	Medicaid Analytic eXtract claims data (2008; N=30,439) for ten states; cross sectional analysis, survival analysis (Cox proportional hazards) to estimate the effect of follow-up services and MAT on time to a behavioral health readmission with censoring at 90 days	Within 14-days after discharge from inpatient hospital care or residential detoxification	Readmissions after inpatient hospital admission for SUD or after residential detoxification	Medication assisted treatment and residential treatment provided after inpatient hospital discharge were associated with reduced risk of 90-day behavioral health readmission (HR= 0.05, p<.001 and HR= 0.61, p<.001 for residential and medication assisted treatment, respectively.
Garner et al. 2010	Adolescents in publicly funded long-term residential treatment for at least 7 days (N=342) were randomized to either a control group that received standard care (SC) or to one of three assertive care conditions (ACC): assertive continuing care, Contingency management, or both.	Within 14-days of discharge from long-term residential care	Substance use and substance use problems as measure by the GAIN substance use scale (SPS). Recovery status measured as abstinence and no dependence symptoms while living in the community during past 30 days prior to follow-up.	Continuity of care was higher for those in the ACC (78%) compared with SC (56%) and continuity of care was a significant predictor of 3-month recovery status (OR= 1.92, p<.05).
Mark et al. 2013	Analyses conducted with MarketScan® Multistate Medicaid Database, 2004-2009, cross	Within 7-days after discharge	Readmissions to treatment with primary SUD diagnosis in days	Hospitals with a larger percentage of patients with a community mental health center (CMHC) follow up visit after discharge

Study	Design	Time After Discharge to Continuity of Care	Outcomes	Results
	section multivariate analyses controlling for hospital and patient level characteristics.		8 to 30 days after discharge	had a lower probability of readmission. A 1 percent increase in a hospital's percent of patients receiving post-discharge continuity of care in a CMHS within 7 days was associated with a 5 percent reduction in the probability of being readmitted.
DeMarce et al. 2008	Randomized controlled study, 150 participants randomized to receive behavioral continuing care adherence intervention or standard treatment. The intervention included meeting with therapist during last week of residential care to develop a behavioral continuing care contract, and meeting again 9 weeks after discharge. The intervention group received attendance prompts, feedback prior to aftercare therapy sessions, certificates for treatment attendance, and AA/NA meetings.	Beginning aftercare after residential treatment and completing 3-, 7-, and 12-month treatment. Beginning aftercare defined as attending at least two treatment sessions per month.	Retention/length of stay in aftercare, abstinence	Continuity of care support after residential treatment was associated with longer stay in aftercare and higher 1-year abstinence rate
Naeger, et al., 2016	Truven Health Analytics MarketScan® Commercial Claims and Encounters data (2010-2014); retrospective multivariate analysis	Within 30 days of discharge	Engagement in at least two unique SUD outpatient treatment visits within 30 days of discharge	17% of patients engaged in SUD treatment within 30 days of hospital discharge. A behavioral health outpatient visit prior to SUD admission increased odds of treatment by 1.34 (CI: 1.25–1.45), an antidepressant prescription drug fill prior to admission increased odds by 1.14 (CI: 1.07–1.21), a benzodiazepine

Study	Design	Time After Discharge to Continuity of Care	Outcomes	Results
				fill prior to admission increased odds by 1.14 (CI: 1.07–1.21), a principal diagnosis for an SUD at index admission increased odds by 2.13 (CI: 1.97–2.30), an alcohol-related disorder diagnosis at index admission increased odds by 3.13 (CI: 2.87–3.42), an additional SUD diagnosis at index admission increased odds by 2.72 (CI: 2.48–2.98).
Ali & Mutter, 2016	Truven Health Analytics MarketScan® Commercial Claims and Encounters data (2010-2014); retrospective analysis	Within 30 days of discharge	Engagement with follow-up services within 30 days following opioid-related hospitalization	40% of patients do not receive any follow-up services within 30 days of opioid-related hospitalization. Only 10.7% received the recommended combination of both medication and therapeutic service, while 43.3% received only therapeutic services and 6.0% received solely SUD medications.
Paddock, et al. 2017	Retrospective cohort study of veterans who received care for SUDs during October 2006-September 2007	Within 30 days of discharge	Morality 12 and 24 months after the end of the observation period	Patients who had two or more diagnosis-related encounters in the inpatient setting within the 30 days of the start of an index episode had an adjusted OR of 0.65 [CI: 0.58, 0.74] and 0.78 [CI: 0.71, 0.85] for 12- and 24-month mortality, respectively.
Schmidt, et al., 2017	Retrospective study of veterans who received SUD detoxification treatment utilizing data from National Patient Care Database (NCPD)	Within 7 days of end date of either patient's detoxification procedure or the end date of a patient's detoxification-	Two year mortality	39.58% of patients who received detoxification services met the 7-day outpatient follow-up after detoxification criteria. Of these patients, 6.73% died in the 2 years after detoxification, while 12.22% of patients who did not meet this criteria had

Study	Design	Time After Discharge to Continuity of Care	Outcomes	Results
		related prescription		died. After adjustment, meeting the numerator criteria was associated with 53.32% lower odds of 2- year mortality (odds ratio [OR] 0.65, [CI] 0.60–0.71, P<0.001).
Liu, et al., 2020	Study of commercial and Medicaid health plans of over 163 million beneficiaries from 2010 to 2016			80% of members who initiated substance use disorder treatment dropped out of treatment after the initial one or two visits.
Englander, et al., 2019	Cohort study using multivariable analysis of Oregon Medicaid claims comparing Improving Addiction Care Team (IMPACT) patients with propensity-matched controls.	Within 34 days of discharge	Engagement in at least two unique SUD outpatient treatment visits within 34 days of discharge	IMPACT patients engaged in SUD treatment following discharge more frequently than controls (38.9% vs. 23.3%, p < 0.01; aOR 2.15, 95% confidence interval [CI] 1.29–3.58). IMPACT participation remained associated with SUD treatment engagement when limiting the sample to people who were not engaged in treatment prior to hospitalization (aOR 2.63; 95% CI 1.46–4.72).
Rubinsky, et al., 2017	Study of patients receiving treatment in Veterans Health Administration (VHA) SUD Residential Rehabilitation Treatment Programs (SUD RRTPs) and Mental Health Residential Rehabilitation Treatment Programs (MH RRTPs) in fiscal year 2012.	Within 7 and 30 days of completion of treatment	Engagement with continuing care within 7 days and 30 days of end of treatment.	Among SUD RRTPs, the mean rate of outpatient SUD/MH continuing care was 59% within 7 days and 80% within 30 days, and the mean rate of SUD continuing care was 63% within 30 days. Among MH RRTPs with a SUD track, these rates were 56%, 75%, and 36%, respectively. There was substantial variability in continuing care rates across the 97 programs: 21%–93% for SUD/MH care within 7 days, 36%–100% for SUD/MH care within 30 days, and 4%–91% for SUD care within 30 days.

Performance Gap

Studies have documented low rates of post discharge follow up after inpatient or residential SUD treatment:

- A claims-based study of Medicaid enrollees ages 18 to 64 who had an inpatient hospital or residential detox admission for SUD found that 33 percent of participants received outpatient or pharmacotherapy treatment within 14 days of discharge (Reif et al., 2017).
- In a study with veterans, 32 percent of patients had two or more continuing care visits during the month after discharge (Schaefer et al., 2005).
- A study of VHA residential treatment programs found that 41 percent of patients participated in two or more outpatient SUD treatment visits during the first month after discharge (Harris et al., 2006).
- Another study found that among 63 Veterans Health Administration (VHA)-funded residential SUD treatment programs, 63 percent of patients engaged in at least one outpatient SUD or mental health visit within 30 days of discharge (Rubinsky et al., 2017).
- In a study of five states' public-sector SUD treatment systems, rates of continuity of care within 14 days of residential discharge ranged from 15 to 60 percent, and continuity rates after an inpatient hospital stay were from 27 to 47 percent (Garnick et al., 2009).
- In a study using private insurance claims data, Smith and Mark (2014), found that 66 percent of patients discharged from inpatient SUD treatment received follow-up SUD outpatient treatment within 30 days of discharge.
- A study using Medicaid claims and public behavioral health agency data found that 27 percent of individuals who were treated for withdrawal management in an inpatient setting received follow-up outpatient treatment within 30 days of discharge (Mark et al., 2006).

Potential Unintended Consequences of the Measure

The implementation of the *Continuity of Care After Receiving Hospital or Residential SUD Treatment* measure may lead providers to avoid treating clients whom they consider less likely or difficult to achieve continuity of care, such as people who are homeless, or to reduce the use of inpatient and residential treatment. These unintended consequences maybe more likely to occur if there are significant financial penalties imposed on providers with low follow-up rates.

Net benefit

Overall, the evidence suggests that the potential benefits of implementing *Continuity of Care After Receiving Hospital or Residential SUD Treatment* measure outweigh the potential costs or unintended consequences. Benefits to Medicaid beneficiaries who have a SUD and receive follow-up within 7 or 30 days following discharge from inpatient or residential SUD treatment include decreased rates of substance use and relapse, fewer readmissions for inpatient treatment, less involvement in criminal justice, and improved employment outcomes. Additionally, the benefits to society include lower costs related to criminal activity and health care. Health care costs for Medicaid could decrease as continuity of care helps sustain beneficiaries' gains from the initial treatment and prevents readmission into high-cost levels of care.

The cost to adopt the measure is relatively low, as the measure is based on claims data. However, facilities will need to take added effort to support increasing rates continuity of care among clients and to ensure the treatment system's capacity to provide this care. In addition, any improvement in the *Continuity of Care After Receiving Hospital or Residential SUD Treatment* measure implies a cost to Medicaid for the continuity of care services. However, the benefits of continuity of care to Medicaid beneficiaries and to society, including those described previously, are greater than the costs of implementing and using this measure.

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

PubMed searches were conducted using keywords: continuity of care, follow-up treatment, residential, inpatient, substance use disorder, treatment, for any type of study since 2000. We focused on the extent to which continuity of care occurs after discharge and the benefits of receiving additional care after leaving inpatient and residential care for SUD.

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

Ali, M. M., & Mutter, R. (2016). Patients Who Are Privately Insured Receive Limited Follow-Up Services After Opioid-Related Hospitalizations. In *The CBHSQ Report*. Substance Abuse and Mental Health Services Administration (US). http://www.ncbi.nlm.nih.gov/books/NBK355361/

Bassuk EL, Hanson J, Greene RN, Richard M, Laudet A. Peer-Delivered Recovery Support Services for Addictions in the United States: A Systematic Review. J Subst Abuse Treat. 2016 Apr;63:1-9. doi: 10.1016/j.jsat.2016.01.003. Epub 2016 Jan 13.

CMS, CMCS Information Bulletin. New Reporting Measures for Substance Use Disorder (SUD)-focused Health Homes, November 17, 2019. https://www.medicaid.gov/federal-policy-guidance/downloads/cib112719.pdf

DeMarce, J. M., Lash, S. J., Stephens, R. S., Grambow, S. C., & Burden, J. L. (2008). Promoting continuing care adherence among substance abusers with co-occurring psychiatric disorders following residential treatment. *Addict Behav, 33*(9), 1104-1112. doi: 10.1016/j.addbeh.2008.02.008

Englander, H., Dobbertin, K., Lind, B. K., Nicolaidis, C., Graven, P., Dorfman, C., & Korthuis, P. T. (2019). Inpatient Addiction Medicine Consultation and Post-Hospital Substance Use Disorder Treatment Engagement: A Propensity-Matched Analysis. *Journal of General Internal Medicine, 34*(12), 2796–2803. https://doi.org/10.1007/s11606-019-05251-9

Garner, B. R., Godley, M. D., Funk, R. R., Lee, M. T., & Garnick, D. W. (2010). The Washington Circle continuity of care performance measure: predictive validity with adolescents discharged from residential treatment. *J Subst Abuse Treat, 38*(1), 3-11. doi: 10.1016/j.jsat.2009.05.008

Garnick, D. W., Lee, M. T., Horgan, C. M., Acevedo, A., & Washington Circle Public Sector, Workgroup. (2009). Adapting Washington Circle performance measures for public sector substance abuse treatment systems. *J Subst Abuse Treat, 36*(3), 265-277. doi: 10.1016/j.jsat.2008.06.008

Harris, A. H., Gupta, S., Bowe, T., Ellerbe, L. S., Phelps, T. E., Rubinsky, A. D., . . . Trafton, J. (2015). Predictive validity of two process-of-care quality measures for residential substance use disorder treatment. *Addict Sci Clin Pract, 10*, 22. doi: 10.1186/s13722-015-0042-5

Harris, A. H., McKellar, J. D., Moos, R. H., Schaefer, J. A., & Cronkite, R. C. (2006). Predictors of engagement in continuing care following residential substance use disorder treatment. *Drug Alcohol Depend, 84*(1), 93-101. doi: 10.1016/j.drugalcdep.2005.12.010

Heslin, K.C., Elixhauser, A., & Steiner, C.A. (2015). Hospitalizations Involving Mental and Substance Use Disorders Among Adults, 2012. *HCUP Statistical Brief #191. June 2015. Agency for Healthcare Research and Quality, Rockville, MD.* Retrieved August 6, 2018, from http://www.hcup-us.ahrq.gov/reports/statbriefs/sb191-Hospitalization-Mental-Substance-Use-Disorders-2012.pdf

Liebschutz JM, Crooks D, Herman D, Anderson B, Tsui J, Meshesha LZ, Dossabhoy S, Stein M. Buprenorphine treatment for hospitalized, opioid-dependent patients: a randomized clinical trial. JAMA Intern Med. 2014 Aug;174(8):1369-76. doi: 10.1001/jamainternmed.2014.2556.

Liu, J., Storfer-Isser, A., Mark, T. L., Oberlander, T., Horgan, C., Garnick, D. W., & Scholle, S. H. (2020). Access to and Engagement in Substance Use Disorder Treatment Over Time. *Psychiatric Services (Washington, D.C.)*, appips201800461. https://doi.org/10.1176/appi.ps.201800461

Manuel JI, Yuan Y, Herman DB, Svikis DS, Nichols O, Palmer E, Deren S. Barriers and facilitators to successful transition from long-term residential substance abuse treatment. J Subst Abuse Treat. 2017 Mar;74:16-22. doi: 10.1016/j.jsat.2016.12.001.

Mark, T. L., Tomic, K. S., Kowlessar, N., Chu, B. C., Vandivort-Warren, R., & Smith, S. (2013). Hospital readmission among medicaid patients with an index hospitalization for mental and/or substance use disorder. *J Behav Health Serv Res, 40*(2), 207-221. doi: 10.1007/s11414-013-9323-5

Mark, T. L., Vandivort-Warren, R., & Montejano, L. B. (2006). Factors affecting detoxification readmission: Analysis of public sector data from three states. *Journal of Substance Abuse Treatment*, *31*(4), 439–445. https://doi.org/10.1016/j.jsat.2006.05.019

McKay, J. R. (2005). Is there a case for extended interventions for alcohol and drug use disorders? *Addiction, 100*(11), 1594-1610. doi: 10.1111/j.1360-0443.2005.01208.x

Naeger, S., Mutter, R., Ali, M. M., Mark, T., & Hughey, L. (2016). Post-Discharge Treatment Engagement Among Patients with an Opioid-Use Disorder. *Journal of Substance Abuse Treatment*, *69*, 64–71. https://doi.org/10.1016/j.jsat.2016.07.004

National Institute on Drug Abuse. Principles of Drug Addiction Treatment: A Research-Based Guide (Third Edition). January, 2018. https://www.drugabuse.gov/publications/principles-drug-addiction-treatment-research-based-guide-third-edition/principles-effective-treatment

Popovici, I., French, M. T., & McKay, J. R. (2008). Economic evaluation of continuing care interventions in the treatment of substance abuse: Recommendations for future research. *Evaluation Review*, *32*(6), 547–568. https://doi.org/10.1177/0193841X08316311

Reif, S., Acevedo, A., Garnick, D. W., & Fullerton, C. A. (2017). Reducing Behavioral Health Inpatient Readmissions for People With Substance Use Disorders: Do Follow-Up Services Matter? *Psychiatr Serv, 68*(8), 810-818. doi: 10.1176/appi.ps.201600339

Rubinsky AD, Ellerbe LS, Gupta S, Phelps TE, Bowe T, Burden JL, Harris AHS. Outpatient continuing care after residential substance use disorder treatment in the US Veterans Health Administration: Facilitators and challenges. Subst Abus. 2018;39(3):322-330. doi: 10.1080/08897077.2017.1391923.

Schmidt, E. M., Gupta, S., Bowe, T., Ellerbe, L. S., Phelps, T. E., Finney, J. W., Humphreys, K., Trafton, J., Vanneman, M. E., & Harris, A. H. S. (2017). Predictive Validity of Outpatient Follow-up After Detoxification as a Quality Measure. *Journal of Addiction Medicine*, *11*(3), 205–210. https://doi.org/10.1097/ADM.0000000000000298

Schaefer, J. A., Ingudomnukul, E., Harris, A. H. S., & Cronkite, R. C. (2005). Continuity of care practices and substance use disorder patients' engagement in continuing care. *Medical Care*, *43*(12), 1234–1241. https://doi.org/10.1097/01.mlr.0000185736.45129.95

Smith, M. W., & Mark, T. L. (2014). Follow-up encounters within 30 days of a substance abuse-related inpatient discharge. *Psychiatric Services (Washington, D.C.)*, *65*(9), 1080. https://doi.org/10.1176/appi.ps.201400198

Stein BD, Kogan JN, Sorbero M. Substance abuse detoxification and residential treatment among Medicaid-enrolled adults: rates and duration of subsequent treatment. Drug Alcohol Depend. 2009 Sep 1;104(1-2):100-6. doi: 10.1016/j.drugalcdep.2009.04.008

Wakeman SE, Metlay JP, Chang Y, Herman GE, Rigotti NA. Inpatient Addiction Consultation for Hospitalized Patients Increases Post-Discharge Abstinence and Reduces Addiction Severity. J Gen Intern Med. 2017 Aug;32(8):909-916. doi: 10.1007/s11606-017-4077-z. Epub 2017 May 19.

Wang SJ, Wade E, Towle J, Hachey T, Rioux J, Samuels O, Bonner C, Kirkpatrick C, O'Loughlin S, Foster K. Effect of Inpatient Medication-Assisted Therapy on Against-Medical-Advice Discharge and Readmission Rates. Am J Med. 2020 May 20:S0002-9343(20)30428-9. doi: 10.1016/j.amjmed.2020.04.025. Epub ahead of print. PMID: 32445720.

Watkins, K. E., Paddock, S. M., Hudson, T. J., Ounpraseuth, S., Schrader, A. M., Hepner, K. A., & Sullivan, G. (2016). Association Between Quality Measures and Mortality in Individuals With Co-Occurring Mental Health and Substance Use Disorders. *Journal of Substance Abuse Treatment*, *69*, 1–8. https://doi.org/10.1016/j.jsat.2016.06.001

1b. Performance Gap

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

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

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

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

Remaining in addiction treatment for an adequate period is critical for recovery (National Institute on Drug Abuse, 2018). Patients often drop out of SUD treatment during transitions from inpatient and residential setting to outpatient settings (Harris, et al. 2006; Naeger et al., 2016; Ali et al., 2016; Reif, et al. , 2017; Rubinsky et al., 2017; Liu et a., 2020). Studies with Medicaid beneficiaries document that, on average, only 25% of them receive a post-discharge follow-up within 14 days of a residential or inpatient SUD stays (although rates vary significantly among states and providers) (Harris et al., 2006; Naeger et al., 2016; Ali et al., 2016; Reif et al., 2017; Rubinsky et al., 2017; Liu et a., 2020).

Research finds that post-discharge follow-up after discharge from an inpatient or residential SUD stay is associated with better outcomes, such as reduced mortality (Harris et al., 2015; Paddock et al., 2017; Schmidt et al., 2017), readmissions (Mark et al.., 2013, Reif et al., 2017), substance use and improved retention in treatment (Demarce et al., 2018, Garner et al., 2010). The use of a performance measure to support post-discharge follow-up after discharge from an inpatient or residential SUD stay can support quality improvement efforts.

NQF endorsed the metric Continuity of care after inpatient or residential treatment for substance use disorder (NQF 3453) for use in Medicaid programs and health plans to stimulate improvement in post-discharge continuity rates. The measure is defined as the percentage of discharges from inpatient or residential treatment for SUD for Medicaid beneficiaries, ages 18–64, followed by a SUD treatment service.

A provider level measure is needed in addition to the Medicaid and health plan level measure for several reasons. First, there is significant variation among providers in post-discharge follow-up after inpatient and residential SUD treatment (Stein et al., 2009; Rubinsky et al., 2018). Creating a provider-level measure allows states, payers, policymakers, and others to target quality improvement to providers that need it. Second, provider-level measures can reveal why some providers have lower follow-up rates than others and identify solutions. Interventions to improve post-discharge follow-up include inpatient addiction consults (Englander et al., 2019), scheduling outpatient appointments before discharge, starting patients on medications to treat opioid use disorder mediations before discharge, using peer navigators, and facilitating obtaining housing and other social supports (Bassuk et al., 2016, Wakeman et al. 2017, Manuel et al., 2017; Liebschutz et al., Wang et al., 2020), Third, NQF SUD measures are being used at the provider-level in Centers for Medicare and Medicaid (CMS) demonstrations, such as in behavioral health home demonstrations; however, they have not been endorsed at the provider level (CMS, 2019). Finally, some states and private health plans are already using the measure at the provider level, for example, New York's Office of Addiction Services and Support is using the measure to help programs improve follow-up rates, and the measures are being reported for New York, Massachusetts, and West Virginia as part of the Shatterproof Atlas portal.

This proposed provider-level measure - Continuity of care after inpatient or residential treatment for substance use disorder - has a similar logic model endorsed at the Medicaid program level. The logic model, and associated evidence, indicates that this measure could help to reduce hospital readmissions (Mark et al., 2013; Reif et al., 2017), decrease substance use and relapse (DeMarce et al., 2008; Garner et al., 2010), and

lower mortality (Harris et al., 2015; Paddock et al., 2017; Schmidt et al., 2017). Potential benefits to society include reduced costs related to lower crime rates and decreased health care expenditures (Popovici, French, & McKay, 2008; Heslin et al., 2015).

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

Performance scores based on Medicaid claims data from 623 providers who treated 81,720 beneficiaries demonstrate a significant performance gap. The median 7-day follow-up rate was 11%, and the median 30-day follow-up rate was 24%. There was also significant variation among providers. Observed scores for the 7-day follow-up ranged from 0% to 99%, with a mean of 19% and SD of 22%. Observed scores for the 30-day follow-up ranged from 0% to 99%, with a mean of 29% and SD of 13%.

Summary Data of Observed Scores

Measure	n	Mean	SD	Min	10th	25th	50th	75th	90th	Max
7 Days	623	19%	22%	0%	0%	4%	11%	25%	52%	99%
30 Days	623	29%	13%	0%	3%	10%	24%	43%	63%	99%

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.

Not applicable.

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

The table below reveals disparities by gender and race in continuity of care after SUD treatment in a hospital or residential setting. The mean describes the mean percentage of patients receiving follow-up within 7 days of discharge from treatment for SUD in a hospital or residential setting by gender or race. Males were less likely to receive follow-up care (15% versus 23%). Blacks were less likely than Whites to receive follow-up care (22% versus 9%).

	Mean	N
Gender		
Male	15%	52,566
Female	23%	29,154

Race	Mean	N
White	22%	44,686
Black	9%	14,853
American Indian/Alaskan Native	26%	887
Asian	11%	489
Hispanic/Latino	18%	1,412
Total	18%	81,720

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

Not applicable.

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

This is not an eMeasure **Attachment:**

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

Attachment**:** Data_Dictionary_for_SUD_Follow_up_Measure.xlsx

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

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

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

Not an instrument-based measure

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

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

Not applicable

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

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

Medicaid discharges, ages 18 to 64, with a principal/primary substance (SUD) diagnosis treated at an inpatient or residential provider that received SUD follow-up treatment within 7 or 30 days after discharge. SUD treatment includes outpatient, intensive outpatient, or partial hospitalization visits; telehealth encounters; or SUD medication fills or administrations; or residential treatment (after an inpatient discharge. Two rates are reported: continuity within 7 and 30 days after discharge.

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

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

For this measure two numerators are calculated (follow-up within 7 days of discharge and follow-up within 30 days of discharge). For the 7-day follow up calculation, the numerator is the total discharges with an outpatient visit, intensive outpatient encounter or partial hospitalization OR telehealth visit with SUD diagnosis in principal position, or filled a prescription for or were administered a medication for SUD within 7 days after discharge. Set this variable equal to 1 if either of the following occur: (a) Follow-up visit or telehealth encounter after index discharge date and on or before index discharge date + 7. SUD diagnosis codes must be in principal position for the follow-up encounter. (b) SUD-related medication fill (see attached Appendix D) on or after index discharge date and on or before index discharge date + 7.

The same process above applies for the 30-day follow-up calculation, but within 30 days after discharge. Set the variable equal to 1 if either of the following occur: (a) Follow-up visit or telehealth encounter after index discharge date and on or before index discharge date + 30. SUD diagnosis codes must be in principal position for the follow-up encounter. (b) SUD-related medication fill (see attached Appendix D) on or after index discharge date and on or before index discharge date + 30.

The measure time period is a calendar year.

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

The denominator are Medicaid beneficiaries, ages 18-64, discharged from inpatient or residential provider with a principal diagnosis of SUD on the inpatient/residential treatment encounter claim.

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 target population for the denominator includes all Medicaid beneficiaries (non-dual eligible) age 18 through 64 years and who had a discharge from SUD inpatient or residential treatment provider with a principal/primary SUD diagnosis during the measurement year which is defined as a calendar year. Eligible discharges are identified based on discharge date.

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

Dual eligible Medicare/Medicaid beneficiaries are excluded. Rationale: Individuals who are covered under Medicare would receive coverage for follow-up treatment medications (e.g. opioid use disorder medications) under Medicare Part D and Medicare Part D claims are not captured in Medicaid claims databases. Therefore follow-up treatment would be missed.

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.)*

Dual eligible (Medicare/Medicaid) beneficiaries (as identified on Medicaid enrollment/beneficiary files)

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.)*

Not applicable.

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = 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.)*

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

Step 1: Identify denominator

Identify Medicaid- only beneficiaries age 18 through 64 years who had a discharge from SUD inpatient or residential treatment with a principal/primary SUD diagnosis during the measurement year. Age is calculated as of December 31st of the measurement year.

Step 1A. Exclude discharge if the date of discharge (for inpatient or residential levels of care) is after December 15 of the measurement year or if the date of discharge is missing.

Step 1B. Exclude discharge if the discharge date occurs on the same day as admission to another inpatient or residential facility. Consider this a transfer; the discharge date from the transfer facility would therefore define the index date.

Step 1C. Exclude any discharges that did not have continuous enrollment with both medical and pharmacy benefits on and within the 30 days of that index discharge date.

Step 2: Identify numerator

Step 2A. Use the Analytic Sample to Create the 7- and 30- day follow-up variables:

a. 7_day_follow-up: Identify discharges with an outpatient visit, intensive outpatient encounter or partial hospitalization OR telehealth visit with SUD diagnosis in principal position, or filled a prescription for or were administered medication for SUD within 7 days after discharge. Set this variable equal to 1 if either of the following occurs:
    i. Follow-up visit or telehealth encounter (Appendix C) after index discharge date and on or before index discharge date + 7. SUD diagnosis codes must be in principal position for the follow-up encounter
    ii. SUD-related medication fill (Appendix D) on or after index discharge date and on or before index discharge date + 7.
b. 30_day_follow-up: Identify discharges with an outpatient visit, intensive outpatient encounter, or partial hospitalization OR telehealth visit with SUD diagnosis in principal position or filled a prescription for or were administered medication for SUD within 30 days after discharge. Set this variable equal to 1 if either of the following occurs:
    i. Follow-up visit or telehealth encounter (Appendix C) after index discharge date and on or before index discharge date + 30. SUD diagnosis codes must be in principal position for the follow-up encounter.
    ii. SUD-related medication fill (Appendix D) on or after index discharge date and on or before index discharge date + 30.

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. The measure is not based on a sample

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. The measure is not based on survey or patient-reported data.

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

If other, please describe in S.18.

Claims, Enrollment Data

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.

The Medicaid Analytic Extract (MAX) files were used to identify discharges from inpatient substance use disorder (SUD) or residential specialty SUD treatment programs with a principal/primary SUD diagnosis on the discharge record (denominator) and the receipt of SUD outpatient or prescription medication treatment within 7 and/or 30 days after discharge (numerator). The Medicaid MAX files used include the following types of files: personal summary (PS), inpatient (IP), other services (OT), long-term care (LT) and drug (RX) files. Data from the PS IP, LT and OT files were used to construct the measure denominator. We used the PS file to limit

the analytic sample based on age and enrollment criteria, and then we used the IP, LT, and OT files to determine whether those beneficiaries met the criteria for the measure denominator. The OT and Rx files enabled us to identify the numerator events (e.g., receipt of SUD outpatient treatment within 7 and/or 30 days after discharge). The PS file contained additional demographic and enrollment information, such as beneficiaries' state, age, sex, and race or ethnicity.

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

SUD_follow_up__testingform__092720.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*)**:**

Measure Title: Continuity of Care After Receiving Hospital or Residential Substance Use Disorder (SUD) Treatment

Date of Submission: **[x/x/2020]**

Type of Measure:

Measure	Measure (continued)
☐ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	*

*cell intentionally left blank

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:	☐ other:

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*).

The Medicaid Analytic Extract (MAX) files were used to identify discharges from hospitals or residential addiction treatment facilities with a primary Substance Use Disorder (SUD) diagnosis on the discharge record (denominator) and the receipt of SUD treatment within 7 and 30 days after discharge (numerator). SUD treatment was defined as an outpatient visit, intensive outpatient encounter, partial hospitalization, or telehealth encounter with a primary SUD diagnosis, or a prescription or administration of a SUD medication.

The Medicaid MAX includes the following file types:

Person summary (PS). Person-level file for Medicaid eligibility and demographic information.

Inpatient (IP). Claim-level file for inpatient hospital stays.

Long-term care (LT). Claim-level file for long-term care institutional stays (nursing facilities, intermediate care facilities for individuals with intellectual disabilities, psychiatric hospitals, and so on).

Other therapy (OT). Claim-level file for a wide variety of services, many of which are provided on an outpatient basis. Most notably, it may contain both residential and other stayover service claims data as claims are assigned to MAX claims file types based upon the category of service provided.

Rx file. Claims-level file provides information on drugs and other services provided by a pharmacy.

Data from the PS IP, LT and OT files were used to construct the measure denominator. We used the IP, LT, and OT files to determine whether those beneficiaries met the criteria for the measure denominator. The OT and Rx files enabled us to identify the numerator events (e.g., receipt of SUD outpatient treatment within 7 and/or 30 days after discharge). The PS file contains demographic and enrollment information, such as beneficiaries' state, disability status, age, sex, and race or ethnicity.

1.3. What are the dates of the data used in testing?

January 1, 2014 – December 31, 2014. The year of data used for testing were based on the most current Medicaid data available at the time that testing began.

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*)

This measure is intended to be reported among all adult Medicaid beneficiaries at the hospital/facility/agency level. A similar measure (NQF#3453) has been developed and endorsed at the population level (Population: State and Medicaid plan).

We identified residential or hospital providers using place of service, HCPC, and revenue codes. After sub-setting to hospital or residential settings, facilities were identified based on the National Provider Identifier (NPI) listed on the medical claims. A patient was attributed to an inpatient or residential provider if the provider submitted a claim for the patient for an encounter with a primary SUD diagnosis. Providers had to have submitted claims for at least 10 patients that met the denominator criteria to be eligible for the measure.

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
x☐ **hospital/facility/agency**	x☐ **hospital/facility/agency**
☐ health plan	☐ health plan
other:	other:

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

A total of 623 providers were included for testing the measure. Of these, 58% were hospitals and 42% were residential facilities.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

The analyses for the 7-day follow-up included a total sample of 81,720 beneficiaries eligible for Medicaid (and not dually eligible for both Medicaid and Medicare) between 18 and 64 years old (including ages 18 and 64) discharged from inpatient or residential facility with SUD as a primary diagnosis in calendar year 2014**.** The analyses for the 30-day follow-up included a total sample of 73,243 eligible beneficiaries (**see table 2)** (Patients without continuous enrollment were excluded)**.** The corresponding identification codes used to identify qualifying encounters and medication fills are listed in appendices A – D.

The most common Medicaid beneficiary category was *Adult* (70.0% and 69.9%, for the 30 day and 7 day follow-up measures, respectively). Over half of the beneficiaries were between the ages of 25-44 (54.9% and 54.5%), were male (63.0% and 64.3%), and were White (55.9% and 54.7%).

Table 2. Beneficiary Characteristics

Measures	Follow-Up 30 Days: Number of beneficiaries (n)	Follow-Up 30 Days: Distribution of beneficiaries (%)	Follow-Up 7 Days: Number of beneficiaries (n)	Follow-Up 7 Days: Distribution of beneficiaries (%)
Total	73,243	100.0	81,720	100.0
Medicaid Category	*	*	*	*
Adult	51,303	70.0	57,114	69.9
Blind/Disabled	17,792	24.3	20,102	24.6
Child	2,509	3.4	2,684	3.3
Other[1]	1,639	2.2	1,820	2.2
Age	*	*	*	*
18-24	9,090	12.4	9,825	12.0
25-44	40,183	54.9	44,512	54.5

[1] The "other" Medicaid category refers to claims with the following corresponding claims codes: 17 (unemployed adult, eligible under Section 1931 of the Act), 3A (individual covered under the Breast and Cervical Cancer Prevention Act of 2000), and ZZ (for months an individual was reported in Medicaid Statistical Information System (MSIS) with a valid T-MSIS eligibility group, but not reported with a MSIS Maintenance Assistance Status (MAS)/MSIS Basis of Eligibility (BOE) assignment).

Measures	Follow-Up 30 Days: Number of beneficiaries (n)	Follow-Up 30 Days: Distribution of beneficiaries (%)	Follow-Up 7 Days: Number of beneficiaries (n)	Follow-Up 7 Days: Distribution of beneficiaries (%)
45-64	23,970	32.7	27,383	33.5
Gender	*	*	*	*
Male	46,117	63.0	52,566	64.3
Female	27,126	37.0	29,154	35.7
Race/ethnicity	*	*	*	*
White	40,959	55.9	44,686	54.7
Black	12,882	17.6	14,853	18.2
American Indian/Alaskan Native	819	1.1	887	1.1
Asian	444	0.6	489	0.6
Hispanic/Latino	1,293	1.8	1,412	1.7
Native Hawaiian/Pacific Islander	71	0.1	77	0.1
Other	7,183	9.8	8,250	10.1
Unknown	9,592	13.1	11,066	13.5

*cell intentionally left blank

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.

The same population was used for all aspects of testing.

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.

Medicaid eligibility is primarily based on income or disability, which are both social risk factors. The other potential social risk factors available in the Max 2014 are racial and ethnic minority status. The measure is not risk adjusted or risk stratified because this is a process measure that applies to all patients who meet the inclusion (denominator) criteria.

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*)

Performance Score Reliability Testing

Reliability testing was conducted to determine whether the measure can distinguish well-performing providers from poorly performing providers. To test reliability, we conducted four types of reliability tests:

1. **Signal-to-noise ratio and effect size tests**: We conducted parametric analysis of variance (ANOVA) and calculated effect size statistics. Both the observed F-ratios and the effect size statistics can be considered measures of signal-to-noise ratios where the signal is the effect created by true differences in underlying provider characteristics and the noise is all variance that is not explained by those differences. Effect size statistics describe the extent to which the independent variable (the provider) influenced the dependent variable (a "success" or "failure" for a patient on a particular measure). For a measure to be reliable, providers should have a demonstrable impact on the quality measure. For each analysis of variance, we calculate the *F* statistic and estimate two effect size statistics: (1) eta squared (η2), which represents the proportion of the variance in the measure that is explained by the provider, and (2) omega squared (ω2), which is similar to η2 but is more robust with regard to small sample size. These tests provide insight into whether the provider has a significant effect on measure performance, as well as the size of that effect. An η2 or ω2 squared value of 0.14 or higher indicates a large effect.
2. **Intra-unit reliability (IUR):** The IUR provides another way to assess a measure's ability to detect true variation. The IUR provides an estimate of the fraction of total variance that is due to signal (i.e., true variation in provider performance) by rescaling the *F* statistic – itself a measure of the ratio of between-groups variance to within-groups variance – using the formula (1-1/F).[2] IUR values are therefore between 0 and 1, with values close to 1 indicating a measure is reliable. An IUR greater than 0.9 is recommended for a measure to be considered reliable.
3. **Adams's rho (ρ):** This approach was developed by Adams[3] and relies on the observed variance of the observed measure score for each as the measure of precision and produces a different reliability statistic for each provider. According to Adams, a 70% reliability demonstrates differences between providers, and a 90% reliability represents statistically significant differences between providers.

[2] Zaslavsky, A. (2001). Statistical issues in reporting quality data: small samples and casemix variation. *International Journal for Quality in Health Care*, vol. 13, 6: 481-488.

[3] Adams, J. L. (2009). The reliability of provider profiling: a tutorial. Accessed from https://www.rand.org/pubs/technical_reports/TR653.html.

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*)

Performance Score Reliability Testing

Signal-to-noise analysis: Results of the signal-to-noise analysis are presented in table 3. The significant *F* statistic at 7-day follow-up ($F = 63.9$, $p < 0.001$) and 30-day follow-up ($F = 56.61$, $p < 0.001$) indicate a significant difference in performance between providers. Eta-squared and omega-squared values 7-day follow-up ($\eta^2 = 0.25$ and $\omega^2 = 0.24$) and 30-day follow-up ($\eta 2 = 0.22$ and $\omega 2 = 0.22$) indicate a large effect size and thus that the differences between provider scores are large.

Table 3: Reliability Results from Signal-to-noise Analysis

Measures	F	η^2	ω^2
7-day follow-up	63.9*	0.25*	0.24*
30-day follow-up	56.61*	0.22*	0.22*

NOTES: *$p < 0.001$, $\eta^2 > 0.14$ indicates a large effect, $\omega^2 > 0.14$ indicates a large effect

Intra-unit reliability (IUR): The calculated IUR was 0.94 and 0.93 for the 7-day and 30-day follow-up, respectively, which exceeded the threshold recommended for determining acceptable reliability.

Adams's rho (ρ): The mean calculated Adams's ρ was 0.94 for 7-day follow-up and 0.93 for 30-day follow-up, which exceeded the 0.7 threshold recommended to indicate acceptable reliability (**Table 4**). For 7-day follow-up, the median ρ was 0.98, the standard deviation was 0.07 and the interquartile range was 0.08. For 30-day follow-up, the median ρ was 0.96, the standard deviation was 0.07 and the interquartile range was 0.10.

Table 4: Reliability Results for Adams's rho (ρ)

Measures	n	Mean	SD	Min	10th	25th	50th	75th	90th	Max
7-day	623	0.94	0.07	0.67	0.83	0.92	0.98	1.00	1.00	1
30-day	623	0.93	0.07	0.71	0.82	0.90	0.96	0.99	1.00	1

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?*)

The reliability testing results indicate that the measure is highly reliable.

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)*

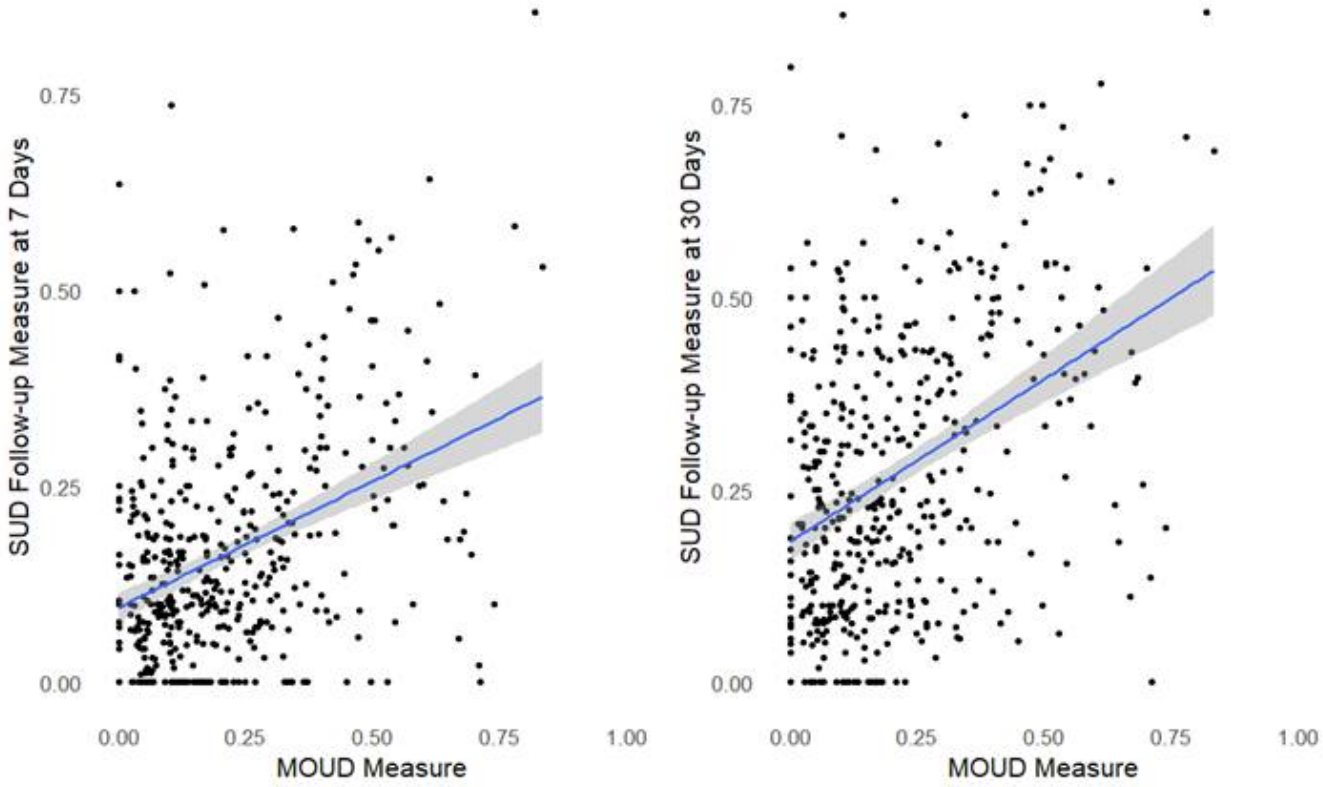
Convergent validity: Convergent validity is established by empirically showing that measures that are conceptually related are in fact statistically correlated to one another. The Continuity of Care After Inpatient or Residential SUD Treatment at the Provider Level (the "SUD Follow-Up" measure) and the Prescription or Administration of Pharmacotherapy for OUD (the "MOUD measure") are conceptually related since both are meant to measure processes that reflect better quality OUD treatment. The MOUD measure captures the percentage of a provider's patients who were Medicaid beneficiaries, ages 18 to 64, with an OUD diagnosis who filled a prescription for, or were administered or ordered, a FDA-approved medication to treat OUD within 30 days of the first attributable OUD treatment encounter with that provider.

To empirically assess the relationship between the two measures, we conducted correlation analysis between the SUD Follow-up measure at 7-day follow-up and the MOUD receipt measure. We calculated the Pearson Product Moment Correlation Coefficient (r), which measures the strength of the association between the two measures. Looking at absolute values, a coefficient value of $r < 0.3$ indicates weak strength, $0.30 \leq r < 0.5$ indicates moderate strength, and $r \geq 0.50$ indicates a strong relationship (Cohen, 2013; Barch, 2019).

1. **2b1.3. What were the statistical results from validity testing**? (*e.g., correlation; t-test*)

Convergent Validity: The measures of MOUD receipt and SUD follow-up had strong, positive, and significant correlations (see **figures 2** below). The correlation between the SUD Follow-up measure at 7-day follow-up and the MOUD measure was r=0.39 (p<0.001), the correlation between the SUD Follow-up measure at 30-day follow-up and the MOUD measure was r=0.39 (p<0.001).

Figure 2. Scatterplot of the SUD Follow-up Measure at 7-Day and 30-Day Follow-Up and the MOUD measure


SUD Follow-up Measure at 7 Days
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
MOUD Measure
SUD Follow-up Measure at 30 Days
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
MOUD Measure

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 correlation coefficient values indicate that there is a significant positive correlation between the SUD Follow-Up measure at 7-Day follow-up and the MOUD receipt. Thus, the convergent validity is supported.

2b2. EXCLUSIONS ANALYSIS

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

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*)

Exclusions and rationale are outlined under section 2b2.3.

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*)

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*)

Dual eligible Medicare/Medicaid beneficiaries are excluded. Rationale: Individuals who are covered under Medicare would receive coverage for follow up treatment medications (e.g. medication assisted treatment) under Medicare Part D and Medicare Part D claims are not captured in Medicaid claims databases. Therefore, follow-up would be missed.

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 risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

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

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

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

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

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

2b3.4a. What were the statistical results of the analyses used to select risk factors?

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

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

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

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

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

As described in 2a.2.2, we conducted signal-to-noise reliability analyses and calculated effect size statistics ((1) eta squared (η2), and (2) omega squared (ω2)), to assess whether the provider has a significant effect on measure performance, as well as the size of that effect. We also calculated the intra-unit reliability (IUR) and Adams's rho (ρ) further assess the measure's ability to detect true variation and statistically significant differences in provider performance.

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*)

As presented in section 2a2.3, the three reliability tests all indicated that the measure can identify statistically and practically meaningful differences in performance across the measure entities.

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 *F*-statistic indicates that the measure scores are significantly different while subsequent statistics indicate reliability, empirically substantiating that the measure can discern clear differences in performance between providers. The measure results therefore demonstrate statistically significant and practically meaningful differences in provider performance and the utility of the measure to foster improvement follow-up up SUD care.

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 non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

This measure is calculated using Medicaid claims data; because submission and completion of claims is tied to provider reimbursement, missing data are rare. Thus, missing data do not have an impact on the measure. Therefore, we did not perform any formal missing data analyses.

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*)

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 non-responders) 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)

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.

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

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

Not applicable.

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

Attachment:

3c. Data Collection Strategy

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

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

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

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)*.

None.

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)
*	Quality Improvement (external benchmarking to organizations) New York, Massachusetts and West Virginia Medicaid https://www.treatmentatlas.org/ Quality Improvement (Internal to the specific organization) New York, Massachusetts and West Virginia Medicaid https://www.treatmentatlas.org/

*cell intentionally left blank

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

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting
- Name of program and sponsor: New York Office of Addiction Supports and Services, Shatterproof ATLAS
- Purpose: Quality improvement
- Geographic area and number and percentage of accountable entities and patients included: New York state (approximately 90 addiction treatment facilities), Shatterproof ATLAS (approximately 130 addiction treatment providers across 3 states, New York, Massachusetts, and West Virginia).
- Level of measurement and setting: Specialty addiction treatment facility

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?*)

Not applicable.

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.*)

See above. The measure is being reported in several states as part of quality improvement efforts.

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.

All Medicaid participating substance use disorder specialty facilities in New York, New York, Massachusetts, and West Virginia

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.

The data are presented in a portal only accessible to providers, state policymakers, and providers. Providers are offered technical assistance material and training to help improve follow-up rates.

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.

The measure was developed with feedback from state Medicaid programs, commercial health plans, addiction treatment providers, patients, families, and other experts. Experts reviewed the measure as part of a NQF sponsored Strategy session. Focus groups were held with providers, patients, and families to obtain feedback on the measures. One Medicaid program and one commercial health plan helped to test and refine the initial specification. The measure was then implemented by three Medicaid programs as part of Shatterproof's Addiction Treatment Locator, Assessment, and Standards (ATLAS) Platform. New York State's Office of Addiction Supports and Services has integrated the measure into its quality improvement activities.

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.

This measure is being considered for initial endorsement. Adoption of this performance measure has the potential to improve the quality of care for Medicaid beneficiaries, who are discharged from inpatient or residential treatment for SUD. Currently the overall rate of continuity of care after inpatient and residential treatment for 7-day follow-up is 17% and 27% for 30-day follow-up, suggesting room for improvement. The Continuity of care after inpatient or residential treatment for substance use disorder measure may be useful

for monitoring the rate of continuing care and encourage states to put interventions in place to increase the rates. This is important because continuity of care (defined in time frames ranging from 7 days to one year post-discharge) has been shown to be related to better outcomes including decreased rates of substance use and relapse (DeMarce, Lash, Stephens, Grambow, & Burden, 2008; McKay & Hiller-Sturmhofel, 2011; Garner et al., 2010), fewer readmissions for inpatient treatment (Mark et al., 2013; Reif et al., 2017), lower risk of death (Harris et al., 2015; Paddock et al., 2017; Schmidt et al, 2017), less involvement in criminal justice (McKay, 2009), and improved employment outcomes (McKay, 2009).

4b2. Unintended Consequences

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

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

There were no unexpected findings identified during testing and early use of this measure in the 3 states.

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

No unexpected benefits were observed during early implementation.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

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

Yes

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

3453 : Continuity of Care after Inpatient or Residential Treatment for Substance Use Disorder (SUD)

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?

No

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

The measure is harmonized with the parallel measure that was developed for use at the health plan or Medicaid program level (NQF 3453). In both measures, the population is Medicaid beneficiaries age 18 – 64. The same diagnosis codes are used to identify substance use disorders. The same services and procedures are included to define follow-up treatment. While NQF# 3453 examines post-discharge follow-up at 7 and 14 days, we are proposing that #3590 Continuity of Care After Receiving Hospital or Residential Substance Use

Disorder (SUD) Treatment be reported at 7 and 30 days. This is because review of the evidence and discussion with addiction professionals supported measuring follow up for an extended period of time. Also, NQF# 3488 Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence, NQF#3489 Follow-up after Emergency Department Visit for MH, and NQF# 0577 Follow-up after a hospitalization for a mental illness indicated that these measures are reported at 7 and 30 days. Finally, HEDIS has implemented NQF #3453 as measuring following-up at 7 and 30 days, not at 7 and 14 days (https://www.ncqa.org/wp-content/uploads/2019/02/20190208_06_FUI.pdf). This difference between using a 14 or 30 day follow-up does not impact interpretability or data collection burden.

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.

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:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): RTI International

Co.2 Point of Contact: Tami, Mark, tmark@rti.org, 240-636-2410-

Co.3 Measure Developer if different from Measure Steward: RTI International

Co.4 Point of Contact: Tami, Mark, tmark@rti.org, 240-636-2410-

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.

An expert panel, supported by NQF, was assembled as a part of a day-long Quality Innovation Measuring Quality of Care in Substance Use Disorder (SUD) Treatment Programs Strategy Session.

The meeting objectives included discussion of considerations for measuring the quality of care for purposes of rating substance use disorder (SUD) treatment programs, gathering feedback on the proposed measure, provision of guidance for adapting the measure for use at the facility-level, and aligning with related measures.

Expert panel members included the following:

Jennifer B. Atkins, MBA

Vice President, Network Solutions, Blue Cross Blue Shield Association

Ellen Bouchery, MS

Principal Program Analyst, Mathematica Policy Research

Teresita Camacho-Gonsalves, PhD, MA

Co-Director of Behavioral Health Team, Human Services Research Institute

Vitka Eisen, EdD, MSW

President & CEO, HealthRight 360

Joseph Lee, MD

Medical Director, Hazelden Betty Ford Foundation Youth Continuum

Miriam Komaromy, MD, FACP, DFASAM

Professor of Medicine, Director of Addiction and Community Health Worker Programs at the ECHO Institute, University of New Mexico Health Sciences Center

Tami Mark, Ph.D., MBA

Senior Director, Behavioral Health Financing and Quality Measurement, RTI International

Tiffany McCaslin, MPP

Senior Policy Analyst, Public Policy, National Business Group on Health

Thomas McLellan, PhD

Founder, Treatment Research Institute

Kirk Moberg, MD, PhD, FASAM, FACP, FAAPL, CPE

Executive Medical Director, UnityPoint Health Illinois Institute for Addiction Recovery

Douglas Nemecek, MD, MBA

Chief Medical Officer – Behavioral Health, and National Medical Officer – Coverage Policy and Trend Review, Cigna

Andre Ostrovsky, MD

Chief Executive Officer, Concerted Care Group

Justin Luke Riley, MBA

President & CEO, Young People in Recovery

Patricia Santora, PhD

Public Health Analyst, Center for Substance Abuse Treatment, Substance Abuse and Mental Health Service Administration (SAMHSA)

Sarah Wattenberg, MSW

Director of Quality and Addiction Services, National Association of Behavioral Healthcare

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2020

Ad.3 Month and Year of most recent revision: 10, 2020

Ad.4 What is your frequency for review/update of this measure? Annual

Ad.5 When is the next scheduled review/update for this measure? 10, 2021

Ad.6 Copyright statement: The Continuity of Care After Receiving Hospital or Residential SUD Treatment measure and specifications were developed by and are owned by RTI International. The measure specifications are not clinical guidelines and do not establish a standard of medical care. RTI makes no representations, warranties, or endorsement about the quality of any organization or provider that uses or reports performance measures and RTI has no liability to anyone who relies on such measures or specifications. RTI holds a copyright for these materials and can alter these materials at any time. These materials may not be modified by anyone other than RTI. Anyone desiring to use or reproduce the materials without modification for a non-commercial purpose may do so without obtaining any approval from RTI. All commercial uses must be approved by RTI and are subject to a license at the discretion of RTI . ©2020 RTI

Limited proprietary coding is contained in the measure specifications for user convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. RTI disclaims all liability for use or accuracy of any coding contained in the specifications.

Some proprietary codes are contained in the measure specifications for convenience of the user. Use of these codes may require permission from the code owner or agreement to a license.

ICD-10 codes are copyrighted © World Health Organization (WHO), Fourth Edition, 2010. CPT © 2010 American Medical Association. CPT is a registered trademark of the American Medical Association. All rights reserved.

Ad.7 Disclaimers: The performance measure, Continuity of Care After Receiving Hospital or Residential SUD Treatment, is not a clinical guideline and does not establish a standard of medical care and has not been tested for all potential applications.

THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND

Ad.8 Additional Information/Comments: RTI Notice of Use. Broad public use and dissemination of these measures, without modification, are encouraged and RTI agrees that all noncommercial uses do not require the consent of the measure developer. Modifications to, and/or commercial use of, a measure requires the prior written consent of RTI and is subject to a license at the discretion of RTI .