NATIONAL QUALITY FORUM
Driving measurable health improvements together

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3312

Corresponding Measures:

Measure Title: Continuity of Care After Medically Managed Withdrawal from Alcohol and/or Drugs

Measure Steward: Centers for Medicare & Medicaid Services

sp.02. Brief Description of Measure: Percentage of discharges from a medically managed withdrawal episode for adult Medicaid Beneficiaries, age 18-64, that was followed by a treatment service for substance use disorder (including the prescription or receipt of a medication to treat a substance use disorder (pharmacotherapy) within 7 or 14 days after discharge). This measure is reported across all medically managed withdrawal settings.

1b.01. Developer Rationale: There is general agreement that continuity of care should occur within a short time after discharge from detox (American Society of Addiction Medicine, 2014). By reporting both a 7- and 14-day follow-up timeframe, the measure balances clinical best practice thinking while recognizing the system's capacity limitations that could cause longer than optimal follow-up timeframes.

---

sp.12. Numerator Statement: Discharges in the denominator who have an inpatient, intensive outpatient, partial hospitalization, outpatient visit, residential, or drug prescription or procedure within 7 or 14 days after discharge from an inpatient hospital, residential addiction program, or ambulatory medically managed withdrawal.

sp.14. Denominator Statement: Adult Medicaid beneficiary discharges from medically managed withdrawal from January 1 to December 15 of the measurement year.

sp.16. Denominator Exclusions: Not applicable: the measure does not have denominator exclusions.

Measure Type: Process

sp.28. Data Source: Claims

sp.07. Level of Analysis: Population: Regional and State; Population: Population

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

Most Recent Endorsement Date: 05/16/2018

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following description for this measure:

- This is a maintenance process measure at the population level of analysis that assesses the continuity of care after medically managed withdrawal from Alcohol and/or Drugs.
- The developer provides a logic model that depicts discharge from a medically managed withdrawal episode leading to the completion of a follow-up visit or prescription within seven or 14 days of discharge, leading to the development of a treatment plan and engagement in treatment, leading to a reduction in readmissions/healthcare costs/criminal justice involvement/improve clinical and employment outcomes.

The developer provides the following evidence for this measure:

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

Summary of prior review in 2016

- The developer provided evidence that supported that continuity of care should occur within a short time after discharge from detoxification. The developer found 11 studies showing association of continuity with a range of better outcomes such as reduction in readmission, less criminal justice involvement, lower mortality, and improved employment

    - The Standing Committee noted that there is strong evidence linking to improved outcomes for individual who receive detoxification services with follow up care. Additionally, the Standing Committee agreed that this measure is important given the current opioid epidemic coupled with high rates of overdose post-detox.
- The Standing Committee requested clarification from the developer regarding the types and timing of pharmacotherapy as it relates to the measure. The developer confirmed that all FDA-approved pharmacotherapies for substance use disorder (SUD) are included in the measure.
- The Standing Committee questioned how the use of monthly treatment and extended release pharmacotherapy, such as naltrexone, might be included in the seven and 14-day timeframes given that the prescription is for 30-days. The developer stated that in their testing they looked at all prescriptions, regardless of the number of days. However, for prescriptions that are given in 30-day dosages, they still require seven or 14 day follow-up given both Substance Abuse and Mental Health Services Administration (SAMHSA) and American Society of Addiction Medicine (ASAM) guidelines.
    - The Standing Committee requested more information on the developer's decision to choose seven- and 14-day follow-up periods. The developer confirmed that the follow-up periods are consistent with SAMHSA and other relevant guidelines. In addition, based on feedback from numerous stakeholders and state agencies, it was suggested that seven days might not be feasible for some organizations, so the developer balanced seven days as clinically appropriate with 14 days as a feasible benchmark for state Medicaid.
- The Standing Committee questioned why telehealth was not included in the measure and the developer confirmed that telehealth had not been an option when the measure was being tested, but agreed that it could be included in future versions of the measure.
- There were concerns from the Standing Committee that same day follow-up visits for newly discharged individuals is not included in the measure. The developer agreed that same day visits are important, but stated that there are limitations in the Medicaid claims data used to calculate the measure making it difficult, if not impossible, to identify same day visits. The Standing Committee hopes to see the inclusion of same day visits in a future iteration of this measure.

Changes to evidence from last review

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

☒ The developer provided updated evidence for this measure:

- Additional publications from 2017 and 2018 were identified as part of the developer's environmental scan and literature review effort. The developer states that no substantive differences in the evidence base were identified and all new evidence aligns with that which was submitted in 2016.
- The evidence submitted now includes mention that patients who are followed up with treatment after detox have lower odds of two-year mortality.
- New evidence also states that substance use disorders have a significant impact on the United States' economy, exceeding $400 billion annually. For every dollar invested in addiction treatment programs such as follow-up after detoxification, between $4 and $7 are directly returned in decreased drug-related crime, criminal justice costs and theft.

Questions for the Committee:

- *The evidence provided by the developer is updated and directionally the same compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?*

Guidance from the Evidence Algorithm

The measure does not assess a health outcome (process measure) (Box 1) -> Systematic review provided (Box 3) -> Evidence is not graded (Box 4) -> The evidence appears to be consistent and high quality -> Rate as MODERATE

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The unit of analysis is at the state level. Data were provided for nine states in calendar year 2018 for both seven-day and 14-day follow-up from detox.
- Seven-day follow-up mean performance rate was 35.7 percent with a standard deviation of 12.29 percent. The lowest state had follow-up visits with 19.96 percent of patients and the highest rate was 52.78 percent. The 25th percentile was 27.85 percent, median was 33.86 percent and the 75th percentile was 47.94 percent.
- 14-day follow-up mean performance rate was 41.01 percent with a standard deviation of 11.78 percent. The lowest state had follow up visits with 24.91 percent of patients and the highest rate was 57.11 percent. The overall population rate at the 25th percentile was 34.50 percent, median was 41.00 percent and the 75th percentile was 52.64 percent.

Disparities

- Results for the 7/14 days follow-up from detox were stratified by age, race/ethnicity, dual eligibility status, gender and Medicaid eligibility category. Results were similar for both measures.
    - Younger patients were more likely than those aged 45-64 years old to have a follow up visit.
    - White non-Hispanic (43.83%/48.91%) and black non-Hispanic patients (44.04%/48.27%) were more likely to have a follow up visit than Hispanic/Latino (35.51%/41.17%) beneficiaries for 7/14 days follow-up, respectively.
    - Dual eligible (25.24%/30.38%) patients were much less likely to have a follow up than Medicaid only (44.70%/49.67%) patients.
    - Males (43.40%/48.58%) were more likely to have a follow up visit than female (41.68%/46.54%) patients.
    - CHIP patients (31.28%/38.00%) were less likely to have a follow up visit than adult (45.68%/50.66%) or disabled (32.73%/37.61%) patients. However, CHIP patients are a very small part of the overall sample, likely reflecting pregnant women and 18-year-olds.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:

1a. Evidence

- evidence for this remains strong
- Agreed that the evidence provided strengthens the measure further and no need for repeat discussion.
- N/A
- Evidence is strong and favorable for continuity of care post discharge from detoxification. Shows a direct relationship to continuity of care and improved outcomes. Not aware of any new studies. There is value to the target population in reduced mortality, improved clinical outcomes.
- Yes. Strong empirical evidence when measure was established. New evidence "includes mention that patients who are followed up with treatment after detox have lower odds of two-year mortality."
- Maintenance process measure-addition of 11 studies; strengthened evidence.
- Generally solid causal ties to important outcomes, with sl updated evidence, directionally the same.
- The logic model behind this process measure i.e. that follow-up is related to ultimate outcomes is reasonable

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

- gap data based only on claims data from 9 states. at the state Medicaid agency level. ave reported at 35.7% for 7 day, 41% for 14 day. percentiles reported: 7 day: 25th%tile: 27.85%, 75th %tile 33.86%; 14 day: 25th % tile 34.5%, 75th%tile: 52.64. Overall low.
- Demonstrated gap that warrants performance measure.
- Current performance data indicate low compliance, despite this measure being in use for many years. If rates have not improved over time, should the committee consider alternatives?
- Follow up rates at the state level show gaps in performance. The population subgroups were provided and there are disparities based on age, race, dual elgibility, gender and insurance.
- Do appear to be gaps in improvement in terms of 7 and 14 day follow up. Disparities evident by age, race/ethnicity, dual/non-dual, and other demographic categories.
- The performance gap is at the individual state level. The evidence does identify the disparities through population subgroups.
- Continues to be a gap.
- There is certainly less than optimal performance and I agree that the gap is moderate

Criteria 2: Scientific Acceptability of Measure Properties

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

2a. Reliability: Specifications and Testing

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

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

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

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

Specifications:

- Measure specifications are mostly clear and precise, however, the previous Standing Committee expressed concerns about the measure reporting both seven-day and 14-day follow-up.

Reliability Testing:

- Reliability testing conducted at the Accountable Entity Level:
    - The developer examined the signal to noise ratio using the beta-binomial model (Adams, 2009). The developer included nine states in the analysis and examined both the seven- and 14-day follow up periods.
        - The average signal to noise ratio for the seven-day follow-up measure was 0.996 and ranged from 0.991 to 0.999 across the states. These results are similar to the findings from the prior 2016 submission but reflect updated CY 2018 claims data.
        - The average signal to noise ratio for the 14-day follow-up measure was .995 and ranged from .990 to .999 across the states. These results are similar to the findings from the prior 2016 submission but reflect updated CY 2018 claims data.
        - The developer states that this high level of reliability can be used to distinguish between states with respect to healthcare quality. The high level of reliability is likely supported by the large sample size at the population/state level.

Questions for the Committee regarding reliability:

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

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

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

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

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

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

Validity Testing

- Validity testing conducted at the Accountable Entity Level:
    - Prior analyses (2016 submission) showed that the odds of subsequent overdose treatment or readmissions between days 15-90 among those with continuity of care within 14 days were 0.917 (95% confidence interval (CI): 0.863, 0.976) as much as those without continuity of care. This translates to an absolute reduction of 1.4 percent and a number needed to treat (NNT) of

71, which was statistically significant (p<.01). Similar results were seen for the follow up within seven days, with an absolute risk reduction of 2.1 percent and an NNT of 48 (p<.01).

- The developer examined validity through examining correlations between 1) NQF 3314 seven-day follow up and NQF 3314 14 day follow up and 2) NQF 3312 with the Healthcare Effectiveness Data and Information Set (HEDIS) Follow-Up (seven and 30 days) After Emergency Department Visit for Alcohol and Other Drug Abuse Dependence.
    - The developer hypothesized that the seven- and 14-day measures would have a large and statistically significant positive correlation with each other and that states performing well on NQF 3312 would also perform well on the select HEDIS measure. Of the nine states included, only six states had data for the HEDIS Medicaid measure.
    - As expected, NQF 3312 seven-day and NQF 14-day follow up had a strong positive correlation (r=0.98, p<.001).
    - NQF 3312 (seven-day) had a weakly positive correlation with seven- and 30-day follow up after emergency department visits for alcohol and other drug abuse dependence (seven/30 days) (r=0.371 and r=0.257 respectively). The developer states that results were not statistically significant likely due to the small sample size (n=6)
    - NQF 3312 (14 day) had a weakly positive correlation with seven- and 30-day follow up after emergency department visits for alcohol and other drug abuse dependence (seven/30 days) (r=0.314 and r=0.143 respectively). The developer states that results were not statistically significant likely due to the small sample size (n=6)

Exclusions

- The measure does not use exclusions.

Risk-Adjustment

- The measure is not risk adjusted or stratified.

Meaningful Differences

- The seven-day follow up measure performance rate ranged from 19.96 percent to 52.78 percent with an average performance of 35.70 percent.
- The 14-day follow up measure performance rate ranged from 24.91 percent to 52.64 percent with an average performance of 41.01 percent.
- A t-test comparing a randomly selected state above the 75th percentile performance and a randomly selected state below the 25th percentile performance was statistically significant (p<.0001) for both the seven- and 14-day follow up rates.
- For the seven-day follow-up rate, eight of the nine states included had a performance of less than 50 percent, while six of the nine states had a performance of less than 50 for the 14-day follow-up rate. None of the states had performance rates above 53 percent and 58 percent for the seven-day and 14-day follow-up rates, respectively. Beneficiaries dually eligible for both Medicare and Medicaid had performance rates that were 19 percentage points lower than beneficiaries eligible for Medicaid only, indicating room for improvement.

Missing Data

- States were selected based on having sufficient data for inclusion in the analyses. They also examined the Transformed Medicaid Statistical Information System (T-MSIS) Research Identifiable Files (RIF) CY18 data to identify states with minimal missing data for the key variables: National Drug Codes (NDCs) and fill dates, dates of service, Medicaid eligibility and type of service.
- From 2016 submission: Some state specific codes were missing in the analysis. However, states implementing the measure will likely have less missing data because they will be able to account for state-specific codes when constructing the measure.

Comparability

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

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*

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

Committee Pre-evaluation Comments:

2a. Reliability

- 2a1. Reliability-Specifications
  - specifications are clear for use with Medicaid claims data
  - I agree about the concerns in reporting both 7 and 14 day follow up. Would opt for 14 day.
  - The numerator includes inpatient stays for SUD following a medically managed withdrawal. If this measure is supposed to assess follow-up care, should inpatient stays for SUD be excluded from the nominator? This may indicate the patient reentered treatment for a relapse, not true follow-up care. Additionally telehealth visits should now be incorporated
  - Data elements clearly defined and descriptors provided. All steps are clear. No concerns about measure being implemented consistently.
  - Reliability testing suggest high level of reliability.
  - No concerns, clearly defined.
  - Reliable
  - Could we ask the developer about any new info/data regarding impact of increasing use of monthly depot forms of naltrexone, visits on same day as discharge, and telephone or televideo visits
- 2a2. Reliability – Testing
  - Given aggregated data at state level for only 9 states, the team is left with beta binomial model to statistically explore capacity to distinguish higher vs. lower performing states.
  - No
  - No
  - No.
  - No
  - No concerns.
  - No concerns

- reliability remains quite good

2b. Validity

- validity explored by examining correlation with adherence rates for two HEDIS measures and found only moderate correlations. The face validity was also explored using a survey item from a "technical expert panel" in 2016. no data presented that adherence is associated with improved outcomes.
- No
- No
- No
- No
- No concerns.
- Reasonable. I would prefer to have a single assessment time (ie, 7 or 14).
- see 5.2a1 above

2b2-2b6. Potential threats to validity

- 2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)
  - risk adjustment may not be feasible with this data source
  - No concerns
  - There is no risk adjustment in the current measure but the developer points out areas where risk adjustment should possibly be considered.
  - No exclusions and not risk adjusted
  - No exclusions, no risk adjustment.
  - The measure does not use exclusions. The measure is not risk adjusted.
  - Reasonable
  - The discrepancies between youth, males, and females are interesting- not sure if it's worth further exploring at this stage
- 2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)
  - Only 9 states were able to report "sufficient data" to be analyzed.
  - No concerns
  - The developer stated that some states have difficulty obtaining data to measure. How much of a concern is this?
  - Yes, there are meaningful differences among states. The 14-day follow up has better rates than the 7-day follow up. Only one set of specifications. Missing data not a threat.
  - No concerns relative to meaningful differences, comparability of performance scores, or missing data/no response.
  - I do not believe the missing data constitutes a threat to the validity.
  - No
  - no

Criterion 3. Feasibility

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

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

- All data elements are available in defined fields in electronic claims and are coded by someone other than the person obtaining the original information.
- The developer used the Data Quality Atlas to identify states with sufficient data quality to be included in testing.
- No applicable fees or licensing required.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:

3. Feasibility

- how can this be feasible if only 9 state Medicaid agency have "sufficient data" ?
- No concerns
- None
- Data come from electronic claims. No issues with the data collection strategy.
- High feasibility, given use of claims data.
- No concerns.
- Feasible
- It's feasible

Criterion 4: Use and Usability

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

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

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

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

Current uses of the measure

Publicly reported? ☐ Yes ☒ No

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

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

Accountability program details

- This measure was previously used in the Medicaid Innovation Accelerator Program (IAP) which ended in 2020. It is now intended for voluntary use by states but is not currently used in an accountability program.

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

Feedback on the measure by those being measured or others

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

***Questions for the Committee**:*

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

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

Improvement results

- For the four states included in the 2016 and 2022 submissions, four-day continuity decreased by an average of 5.7 percentage points and seven-day continuity decreased by an average of 4.0 percentage points. For all nine states included in this submission, the 14-day performance rate is 47.30 percent and the seven-day performance rate is 42.32 percent, showing improvement from previous testing results of 36.5 percent continuity within 14 days and 28.8 percent within seven days.

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

Unexpected findings (positive or negative) during implementation

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

Potential harms

- No harms have been identified.

***Questions for the Committee**:*

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

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

Committee Pre-evaluation Comments:

4a. Use

- not publicly report, not used by accountability program, no feedback on measure performance
- No concern
- The measure is for voluntary use only at this time. Only 9 states submitted data in the most recent submission. If the rates are showing little improvement and the measure is not being widely used, is there a need for the measure?
- Not being publicly reported. Results are not available outside of the groups being measured. States can voluntarily use the measure. Users of the measure can provide feedback, it is incorporated when changes are made, and users are provided results and given assistance in interpretation.
- Data are neither publicly reported nor used in any current/future accountability programs. May be more of a future opportunity for further vetting.
- It is intended for voluntary use by states and is not part of an accountability program. The opportunity to provide feedback was given.
- Not really being used widely, not clear if feedback was solicited
- It's a significant concern that the use has declined since the demonstration in the Medicaid Innovation Accelerator Program and that currently it's not used at all for accountabilty. Big Red Flag!

4b. Usability

- only 9 state Medicaid agencies able to report enough data. With mandatory reporting of BH core measures in 2024, did CMS not invest in trying to get this one endorsed as a maintenance measure?
- Agreed, no unexpected findings or harms.
- N/A
- There have been improvements in resting results over time. When people are provided with continuity of care, there is a greater chance for improved outcomes and reduced mortality. No unexpected findings and no harms.
- Don't appear to be any unintended consequences. Would be interesting to better understand any information about why continuity may have decreased in the 4 states from 2016 to 2022 - perhaps impacts related to COVID-19. That said, the 9 states overall did see improvement.
- The benefits are high for quality care and there are no harms.
- has been shown to promote improvement
- This is an important measure and the fact that it's no longer being used suggests we need to discuss why, whether any modifications might kick start use, or whether it should continue. It does not improve care if it's never used.

Criterion 5: Related and Competing Measures

Related measures

- NQF #0004 Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment
- NQF #2605 Follow-Up After Emergency Department Visit for Mental Illness or Alcohol and Other Drug Abuse or Dependence
- NQF #3453 Continuity of Care after Inpatient or Residential Treatment for Substance Use Disorder (SUD)

Harmonization

- Follow-up time period: NQF #2605 examines follow-up care seven days and 30 days after discharge. Similar to NQF #3312, NQF #3453 examines continuity of care seven and 14 days after discharge from inpatient or residential treatment for SUD. NQF #3312 focuses on continuity of care seven days and 14 days after medically managed withdrawal. The 14-day follow-up time period aligns with NQF #0004, NQF #3453 and the non-NQF endorsed Continuity of Care After Detoxification measure developed by the Washington Circle, and reflects the input of some public commenters that adults should receive some type of care within two weeks of discharge from detoxification.
- Place of Service: NQF #3312 differs from the other related measures by examining continuity of care after medically managed withdrawal by place of service. NQF #2605 focuses on emergency department visit follow-up while NQF 0004 includes initiation of treatment through an inpatient AOD admission, outpatient visit, intensive outpatient encounter or partial hospitalization, telehealth, or medication treatment. NQF #3453 assesses follow-up after discharges from inpatient or residential treatment for substance use disorder.
- Diagnoses: NQF #2605 requires a primary diagnosis of alcohol and other drug dependence (AOD) for the follow-up service. NQF #3453 numerator for follow-up requires a prescription for, administered, or ordered a medication for SUD. The denominator for NQF #3453 requires a discharge from inpatient or residential treatment for SUD with a principal SUD diagnosis. NQF #3312 requires a primary or secondary diagnosis of AOD. We allow a primary or secondary AOD diagnosis to address potential inaccuracies in how AOD diagnoses are coded. For example, some providers may be concerned about the stigma associated with an AOD diagnosis and therefore code it as a secondary diagnosis. Also, for adults with co-occurring mental health and AOD disorders, the assignment of primary and secondary diagnoses can be challenging and sometimes arbitrary.
- The differences in follow-up time period, location, and diagnoses between NQF #2605, NQF #3453 and NQF #3312 do not impact the measure's interpretability in which a higher rate is indicative of better quality. These measures rely on administrative data. The differences in measure specifications between #2605, #3453 and #3312 are minor and expected to have minimal impact on data collection burden.

Committee Pre-evaluation Comments:

5: Related and Competing Measures

- N/A
- All related measures have been harmonized.
- Three related measures, but harmonization appears adequate for all: NQF #0004 Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment; NQF #2605 Follow-Up After Emergency Department Visit for Mental Illness or Alcohol and Other Drug Abuse or Dependence; and NQF #3453 Continuity of Care after Inpatient or Residential Treatment for Substance Use Disorder (SUD)

- No competing measures; related measures NQF #0004 Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment ● NQF #2605 Follow-Up After Emergency Department Visit for Mental Illness or Alcohol and Other Drug Abuse or Dependence ● NQF #3453 Continuity of Care after Inpatient or Residential Treatment for Substance Use Disorder (SUD)
- Not significantly.
- It would be helpful if someone could research the data from NQF 0004, 2605, and 3453 to see compare and contrast their use rates and improvement rates with each other and this measure.

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

Member Expression of Support

- No public comments received.

Scientific Acceptability Evaluation

RELIABILITY: SPECIFICATIONS

1. **Have measure specifications changed since the last review? ☐ Yes ☒ No**
2. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☐ No**
3. **Briefly summarize any changes to the measure specifications and/or concerns about the measure specifications.**
   - [Summary]

RELIABILITY: TESTING

4. **Did the developer conduct new reliability testing? ☒ Yes ☐ No**

   4a. If no, summarize the Standing Committee's previous feedback:
   - [Summary]

   4b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:
   - In 2016, the developer used data from Medicaid Analytic eXtract (MAX) 2013 and 2014 eligible, inpatient, other services, long-term care, and drug files. In 2021, the developer used data from the Transformed Medicaid Statistical Information System (T-MSIS) which contains beneficiary, service utilization, administrative claims, and expenditure data from the Medicaid population, including both fee-for-service and managed care payers.
   - In both 2021 and 2016, the developer used a beta-binomial model to indicate signal-to-noise.
   - 2016 testing found that the average reliability score for both seven-day and 14-day follow-up was 0.99 (developer states that SUD-5 refers to NQF #3312, the measure under consideration).
5. **Reliability testing level: ☒ Accountable-Entity Level ☐ Patient/Encounter Level ☐ Neither**
6. **Reliability testing was conducted with the data source and level of analysis indicated for this measure:**

   ☒ Yes ☐ No
7. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☐ Yes ☐ No
8. **Assess the method(s) used for reliability testing:**

Testing attachment, section 2a2.2

- The developer conducted Accountable Entity Level testing using a signal-to-noise analysis. The signal-to-noise analysis was conducted for the 9 states overall as well as separately for each state for both the seven-day and 14-day continuity measures.
- Updated testing was provided for the spring 2022 submission. Testing was conducted using the Transformed Medicaid Statistical Information System (T-MSIS), which contains beneficiary, service utilization, administrative claims, and expenditure data for the Medicaid population, including those covered through both fee-for-service (FFS) and managed care payers.
    - The measurement period to identify events for the denominator was January 1, 2018 - December 15, 2018.
    - Data was representative of nine states across the United States (U.S.): one Northeastern state, two Midwestern states, four Southern states, two Western states. The states were blinded to protect confidentiality. 52,351 beneficiaries with at least one detoxification episode during the year were identified from this testing cohort. (*Link to Table 1.)*

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

   Testing attachment, section 2a2.3

    - Average signal-to-noise reliability was 0.996 for the seven-day continuity rate and ranged from 0.991 to 0.999; for the 14-day continuity rate it was 0.995, ranging from 0.990 to 0.999.
        - The developer states that this represents high signal-to-noise reliability.
        - The developer also assumes that the finding of high reliability is supported by large sample sizes at the state level. The average number medically managed withdrawal episodes was 8,425 (ranging from 2,545 to 30,157), and the average number of beneficiaries receiving continuity of care within 14 days of discharge from a medically managed withdrawal program was 3,986 (ranging from 634 to 15,875).
    - In their 2016 submission, the developer found that the average reliability score for both the seven- and 14-day continuity rates was 0.98 across states with a range from 0.98 to 0.99.
        - For the seven-day alone average reliability was 0.99 (range of 0.99-0.99) and for the 14-day alone average reliability was 0.99 (range of 0.98-0.99).
        - In 2016 the developer also used a Spearman rank correlation to compare consecutive pairs of quarters (Q1 vs. Q2, etc.) and demonstrated that the measure was consistent across the measurement year for both continuity rates ($p<0.001$ for all pairwise correlation coefficients).

10. 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)

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

    ☐ **Yes** ☐ **No** ☒ **Not applicable** (patient/encounter level testing was not performed)

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

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

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

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

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

13. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
    - Specifications are precise (Box 1) -> Empirical testing using the measure as specified (Box 2) -> Testing conducted at the accountable entity level (Box 4) -> Testing method is appropriate (signal-to-noise) (Box 5) -> Based on the reliability statistic and scope of testing, there is high certainty the measure is reliable (Box 6a) -> Rate as HIGH

VALIDITY: TESTING

14. **Did the developer conduct new validity testing? ☒ Yes ☐ No**

    14a. If no, summarize the Standing Committee's previous feedback:
    - [Summary]

    14b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:
    - In 2016, the developer conducted convergent validity testing by examining the association between the presence/absence of continuity of care (the construct of the measure, defined as having a follow-up visit within seven or 14 days after discharge from detox) and presence/absence of a subsequent overdose treatment or detox readmission.
        - The odds of subsequent overdose treatment or readmissions between days 15-90 among those with continuity of care within 14 days were 0.917 (95% confidence interval (0.863, 0.976).
    - In 2021, the developer conducted convergent validity testing by examining the measure's relationship to a HEDIS measure, *Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence* (FUA), as well as whether the seven- and 14-day rates were correlated with each other, hypothesizing positive relationships for both. Details are provided below.
15. **Validity testing level (check all that apply):**

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

    NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.
16. **If patient/encounter level validity testing was provided, was the method described and appropriate for assessing the accuracy of ALL critical data elements? NOTE:** Data element validation from the literature is acceptable.

    ☐ **Yes**

    ☐ **No**

    ☒ **Not applicable** (patient/encounter level testing was not performed)
17. **Method of establishing validity at the accountable-entity level:**

    ☐ **Face validity**

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

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

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

☒ **Yes**

☐ **No**

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

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

Submission document: Testing attachment, section 2b2.2

- As with reliability, testing was conducted using T-MSIS data from January 1, 2018 - December 15, 2018. For the validity analysis, the team calculated the Spearman correlation between NQF #3312 and related measures for the behavioral health population in the Healthcare Effectiveness Data and Information Set (HEDIS) from measure year 2018. (Spearman was calculate rather than Pearson correlations due to the number of states included in the analyses.)
    - The developer tested the correlation of NQF #3312 with the HEDIS *Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence* (FUA) measure. They hypothesized this would be a positive correlation since the two measures address similar care coordination processes and access to care for shared populations
    - The developer also tested whether the seven- and 14-day indicators for NQF #3312 were correlated with one another. They hypothesized this would be a large and statistically significant positive correlation.

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

Submission document: Testing attachment, section 2b2.3

- The seven- and 14-day continuity rates for NQF #3312 had a strong positive correlation (r=0.98, p<0.001).
- The 14-day follow-up had a weak positive correlation with FUA (FUA seven-day follow-up: r=0.31, p=0.54; 30-day follow-up: r=0.14, p=0.78). The seven-day follow-up also had a weak positive correlation with FUA (FUA seven-day follow-up: r=0.37, p=0.47; 30-day follow-up: r=0.26, p=0.62).
    - The developer asserts that convergent validity with FUA was not statistically significant at P<0.05 given the smaller sample size. Of the 9 states used for measure testing, only 6 states had HEDIS FUA Medicaid data available.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

- The measure does not use exclusions.

22. **Risk Adjustment**

22a. **Risk-adjustment method**

☒ None (only answer Question 20b and 20e) ☐ Statistical model ☐ Stratification

☐ Other method assessing risk factors (please specify)

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

☐ Yes ☐ No ☒ Not applicable

22c. **Social risk adjustment:**

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

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

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

22d.**Risk adjustment summary:**

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

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

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

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

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

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

- [Summary]

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

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

- The seven-day follow up measure performance rate ranged from 19.96 percent to 52.78 percent with an average performance of 35.70 percent.
- The 14-day follow up measure performance rate ranged from 24.91 percent to 52.64 percent with an average performance of 41.01 percent.
- A t-test comparing a randomly selected state above the 75th percentile performance and a randomly selected state below the 25th percentile performance was statistically significant (p<.0001) for both the seven- and 14-day follow up rates.
- For the seven-day follow-up rate, eight of the nine states included had a performance of less than 50 percent, while six of the nine states had a performance of less than 50 for the 14-day follow-up rate. None of the states had performance rates above 53 percent and 58 percent for the seven-day and 14-day follow-up rates, respectively. Beneficiaries dually eligible for both Medicare and Medicaid had performance rates that were 19 percentage points lower than beneficiaries eligible for Medicaid only, indicating room for improvement.

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

- Not applicable.

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

- States were selected based on having sufficient data for inclusion in the analyses. They also examined the T-MSIS RIF CY18 data to identify states with minimal missing data for the key variables: National Drug Codes (NDCs) and fill dates, dates of service, Medicaid eligibility and type of service.
- From 2016 submission: Some state specific codes were missing in the analysis. However, states implementing the measure will likely have less missing data because they will be able to account for state-specific codes when constructing the measure.

For cost/resource use measures ONLY:

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

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

☐ Yes ☐ Somewhat ☐ No (If "Somewhat" or "No", please explain)

27. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**
28. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

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

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

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

    ☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the accountable-entity level and the patient/encounter level is required; if not conducted, should rate as INSUFFICIENT.)
29. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

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

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**
31. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**
    - [Summary]

ADDITIONAL RECOMMENDATIONS

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

Developer Submission

Criteria 1: Importance to Measure and Report

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

---

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

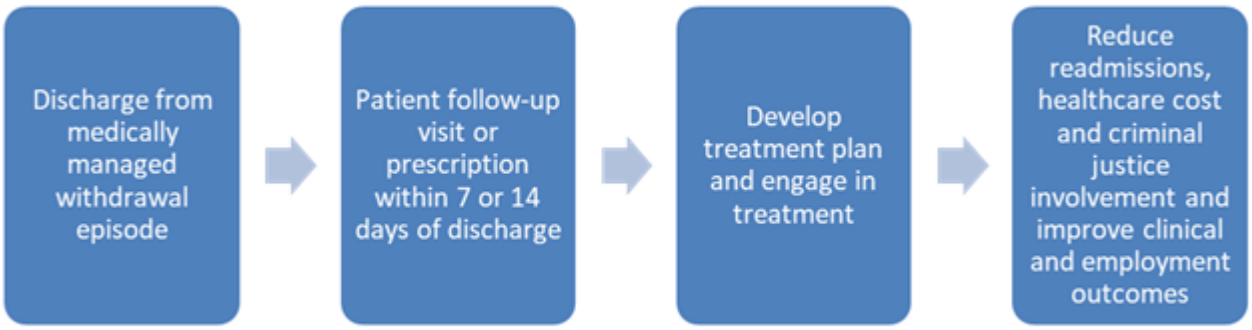
[Response Begins]
No
[Response Ends]

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:
2021 Submission:
Updated evidence information here.
2018 Submission:
Evidence from the previous submission here.

1a.01. Provide a logic model.

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

[Response Begins]
2022 Submission

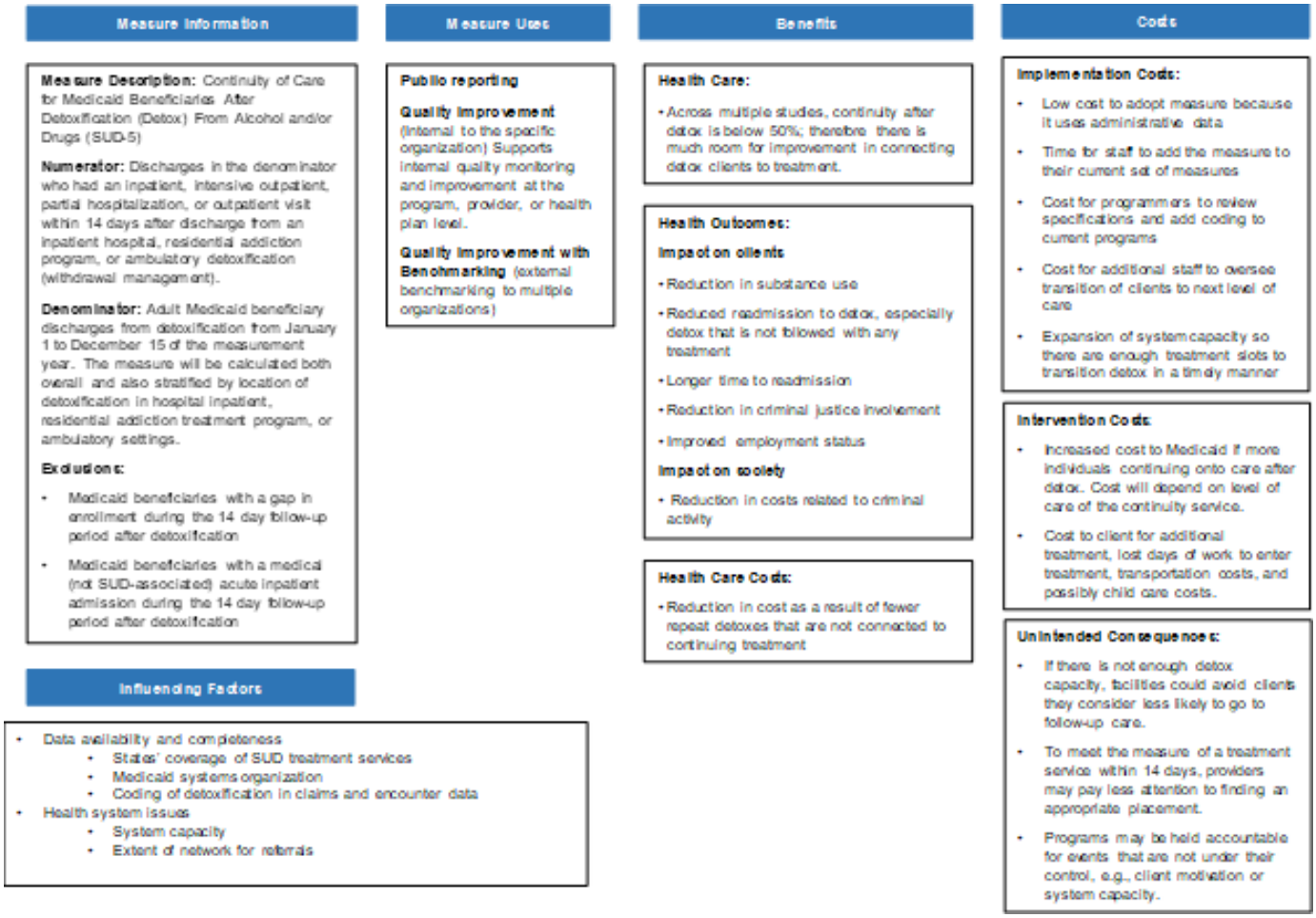


2016 Submission
Continuity of care after detox has been low across multiple sites, and therefore, endorsement and implementation of a measure may be used for public reporting to drive quality improvement. The potential benefits of such a measure at the client level include reduction in substance use (McCusker, Bigelow, Luippold, Zorn, & Lewis, 1995; McLellan, Weinstein,

Shen, Kendig, & Levine, 2005), readmission to detox (Carrier et al., 2011; Ford & Zarate, 2010; Lee et al., 2014; Mark, Vandivort-Warren, & Montejano, 2006), and criminal justice involvement (Ford & Zarate, 2010). The impact on clients also includes longer time to readmission and improved employment status (Ford & Zarate, 2010). The benefits to society include a reduction in costs related to criminal activity, and a reduction in healthcare costs as a result of fewer repeat detoxes (Alexandre et al., 2012; Kertesz, Horton, Friedmann, Saitz, & Samet, 2003; McCollister & French, 2003; McCollister, French, & Fang, 2010).

Factors that influence the measure are data availability and completeness and systems limitations. The former relates to the quality of data that reporting entities have available to them for calculating the measure, while the latter relates to system capacity and whether there is an adequate number of treatment slots to transition patients to as they are discharged from detox.

The logic model below also shows costs of implementing and costs of maintaining the measure as well as potential unintended consequences.



[Response Ends]

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

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

[Response Begins]

Other (specify)

[Other (specify) Please Explain]

2022 Submission

Not applicable. Evidence is not based on a systematic review.

[Response Ends]

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

Evidence - Systematic Reviews Table (Repeatable)
Group 1 - Evidence - Systematic Reviews Table

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A

[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
N/A
[Response Ends]

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

[Response Begins]
2022 Submission
This measure was developed as a result of an extensive environmental scan. In 2016, the team performed an environmental scan of existing SUD quality measures and those under development, identified the major gaps in SUD quality measurement, and recommended measure concepts and domains to fill these gaps through development and testing of de novo or adapted measures. The scan included a targeted literature review and interviews with key stakeholders representing states, managed care organizations, researchers, consumers, and providers, to identify the most promising and meaningful measures for the Medicaid program.
Throughout the process, we were guided by the priorities outlined in the SAMHSA National Behavioral Health Quality Framework (NBHQF). The NBHQF goals reflect an effort to harmonize and prioritize measures that reflect the core principles of SAMHSA, as well as support the CMS National Quality Strategy.
2016 Submission
This measure is proposed as a result of an extensive environmental scan. We performed an environmental scan of existing SUD quality measures and those under development, identified the major gaps in SUD quality measurement, and recommended measure concepts and domains to fill these gaps through development and testing of de novo or adapted measures. The scan included a targeted literature review and interviews with key stakeholders representing states, managed care organizations, researchers, consumers, and providers, to identify the most promising and meaningful measures for the Medicaid program.
Throughout the process, we were guided by the priorities outlined in the SAMHSA National Behavioral Health Quality Framework (NBHQF). The NBHQF goals reflect an effort to harmonize and prioritize measures that reflect the core principles of SAMHSA, as well as support the CMS National Quality Strategy.
SAMHSA (2015a). National Behavioral Health Quality Framework. Retrieved December 9, 2015, from https://www.SAMHSA.gov

[Response Ends]

1a.14. Briefly synthesize the evidence that supports the measure.
[Response Begins]
2022 Submission

Not getting patients into treatment after detox is a missed opportunity to connect them to the treatment system and studies show that having continuity of care after detox is associated with the better outcomes of longer time to repeat detox and fewer detox readmissions. Research supports the need for individuals with SUD to not only receive timely follow-up care following medically managed withdrawal, but also to stabilize or cease using the substance(s) and engage in ongoing treatment to prevent relapse (NIDA, 2018a; Proctor & Herschman, 2014). Individuals who receive timely follow-up care may be more likely to complete treatment or receive more days of treatment than those who leave care prematurely (Proctor & Herschman, 2014). Patients who entered treatment within three days of inpatient detox discharge were less likely to have repeat crisis detox visits than those who did not enter treatment (Carrier et al., 2011). Another study (Lee et al., 2014), using administrative data from five states found that patients who entered treatment with 14-days of detox discharge were less likely to be readmitted to detox and that continuity of care was particularly effective in reducing readmissions to another detox that was not followed with treatment. Adults who receive at least one treatment service within 14 days of detox and two additional services in the next 30 days have a lower hazard of having a detoxification admission over a year (Acevedo et al., 2015).
Improved employment status, reduced criminal activity, and longer periods of abstinence have also been found in patients who had continuity of care after detox (Ford & Zarate, 2010). Those who entered treatment after detox were over four times as likely to being employed within three months of discharge as well as fewer arrests and fewer days in jail. Furthermore, patients who are followed up with treatment after detox have lower odds of 2-year mortality (Schmidt et al., 2017). In spite of the support that continuity of care within a short window of time after leaving detox is related to better outcomes, many do not continue onto care (Carrier et al., 2011; Acevedo et al., 2015).
Additionally, substance use disorders have a significant impact on the United States' economy, exceeding $400 billion annually (NIDA, 2017). For every dollar invested in addiction treatment programs such as follow-up after detoxification, between $4 and $7 are directly returned in decreased drug-related crime, criminal justice costs and theft (NIDA, 2018b).
2016 Submission
Not getting patients into treatment after detox is a missed opportunity to connect them to the treatment system and studies show that having continuity of care after detox is associated with the better outcomes of longer time to repeat detox and fewer detox readmissions. Individuals who experience detoxification not followed by rehabilitative treatment are likely to relapse to substance use, which may result in readmission to another detox (McLellan et al., 2005). Patients who entered treatment within three days of inpatient detox discharge were less likely to have repeat crisis detox visits than those who did not enter treatment (Carrier et al., 2011). Another study (Lee et al., 2014) using administrative data from five states found that patients who entered treatment with 14-days of detox discharge were less likely to be readmitted to detox and that continuity of care was particularly effective in reducing readmissions to another detox that was not followed with treatment. A longer period of time between detox admissions is generally viewed as a better outcome, since it indicates that the individual is experiencing a longer period before a relapse occurs. Several studies have reported that time to readmission was longer when the client continued to treatment after detox (Mark et al., 2006; Thakur, Hoff, Druss, & Catalanotto, 1998).
Improved employment status, reduced criminal activity, and longer periods of abstinence have also been found in patients who had continuity of care after detox (Ford & Zarate, 2010). Those who entered treatment after detox were over four times as likely to being employed within three months of discharge as well as fewer arrests and fewer days in jail. Furthermore, patients who are followed up with treatment after detox have lower rates of drug use at follow-up (McCusker et al., 1995). In spite of the support that continuity of care within a short window of time after leaving detox is related to better outcomes, many do not continue onto care (Campbell et al., 2010; Carrier et al., 2011; Carroll, Triplett, & Mondimore, 2009; Mark, Dilonardo, Chalk, & Coffey, 2003; Mark et al., 2006; Stein, Kogan, & Sorbero, 2009).

[Response Ends]

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

[Response Begins]
2022 Submission
PubMed searches were conducted using keywords: detox, withdrawal management, continuity of care, and outcomes. In addition, titles of key articles or author names were entered into Google Scholar to identify related articles.
In addition, a technical expert panel (TEP) was convened to discuss development of this measure on April 18, 2016, April 20, 2016, and September 21, 2016. The members of the panel advised as to which measures, from a list of measures that were found through an environmental scan as filling a measurement gap, they perceived to be important to develop and test. The TEP rated the continuity of care after detox measure high priority for development.
2016 Submission

PubMed searches were conducted using keywords: detox, withdrawal management, continuity of care, and outcomes. In addition, titles of key articles or author names were entered into Google Scholar to identify related articles.
In addition, a technical expert panel (TEP) was convened to discuss development of this measure on April 18, 2016, April 20, 2016, and September 21, 2016. The members of the panel advised as to which measures, from a list of measures that were found through an environmental scan as filling a measurement gap, they perceived to be important to develop and test. The TEP rated the continuity of care after detox measure high priority for development.

[Response Ends]

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

[Response Begins]
2022 Submission

Alexandre, P. K., Beulaygue, I. C., French, M. T., McCollister, K. E., Popovici, I., & Sayed, B. A. (2012). The economic cost of substance abuse treatment in the state of Florida. *Eval Rev, 36*(3), 167-185. doi: 10.1177/0193841X12450164
Acevedo, A., Garnick, D., Ritter, G., Lundgren, L., & Horgan, C. (2016). Admissions to detoxification after treatment: Does engagement make a difference? *Substance abuse*, 37(2), 364–371. https://doi.org/10.1080/08897077.2015.1080784
Campbell, B. K., Tillotson, C. J., Choi, D., Bryant, K., DiCenzo, J., Provost, S. E., . . . McCarty, D. (2010). Predicting outpatient treatment entry following detoxification for injection drug use: the impact of patient and program factors. *J Subst Abuse Treat, 38 Suppl 1*, S87-96. doi: S0740-5472(10)00024-3 [pii]10.1016/j.jsat.2009.12.012
Carrier, E., McNeely, J., Lobach, I., Tay, S., Gourevitch, M. N., & Raven, M. C. (2011). Factors associated with frequent utilization of crisis substance use detoxification services. *J Addict Dis, 30*(2), 116-122. doi: 936254277 [pii]10.1080/10550887.2011.554776
Ford, L, & Zarate, P. (2010). Closing the gaps: The impact of inpatient detoxification and continuity of care on client outcomes. *Journal of Psychoactive Drugs, SARC Supplement 6, September*, 303-314.
Lee, M. T., Horgan, C. M., Garnick, D. W., Acevedo, A., Panas, L., Ritter, G. A., . . . Reynolds, M. (2014). A performance measure for continuity of care after detoxification: relationship with outcomes. *J Subst Abuse Treat, 47*(2), 130-139. doi: 10.1016/j.jsat.2014.04.002
McCollister, K. E., French, M. T., & Fang, H. (2010). The cost of crime to society: new crime-specific estimates for policy and program evaluation. *Drug Alcohol Depend, 108*(1-2), 98-109. doi: 10.1016/j.drugalcdep.2009.12.002
National Institute on Drug Abuse (NIDA). (2018a). Drugs, Brains, and Behavior: The Science of Addiction. National Institute on Drug Abuse, July 2018. Retrieved from https://www.drugabuse.gov/publications/drugs-brains-behavior-science-addiction/treatment-recovery.
National Institute on Drug Abuse (NIDA). (2018b). Principles of Drug Addiction Treatment: A Research-Based Guide (Third Edition). National Institute on Drug Abuse, 17 Jan. 2018. Retrieved from https://www.drugabuse.gov/publications/principles-drug-addiction-treatment-research-based-guide-third-edition
National Institute on Drug Abuse (NIDA). (2017). Trends & Statistics. National Institute on Drug Abuse, 24 Apr. 2017. Retrieved from https://www.drugabuse.gov/related-topics/trends-statistics#supplemental-references-for-economic-costs.
Proctor, S. L., & Herschman, P. L. (2014). The Continuing Care Model of Substance Use Treatment: What Works, and When Is "Enough," "Enough?" *Psychiatry Journal*, 2014, e692423. https://doi.org/10.1155/2014/692423
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

2016 Submission

Alexandre, P. K., Beulaygue, I. C., French, M. T., McCollister, K. E., Popovici, I., & Sayed, B. A. (2012). The economic cost of substance abuse treatment in the state of Florida. *Eval Rev, 36*(3), 167-185. doi: 10.1177/0193841X12450164
Campbell, B. K., Tillotson, C. J., Choi, D., Bryant, K., DiCenzo, J., Provost, S. E., . . . McCarty, D. (2010). Predicting outpatient treatment entry following detoxification for injection drug use: the impact of patient and program factors. *J Subst Abuse Treat, 38 Suppl 1*, S87-96. doi: S0740-5472(10)00024-3 [pii]10.1016/j.jsat.2009.12.012
Carrier, E., McNeely, J., Lobach, I., Tay, S., Gourevitch, M. N., & Raven, M. C. (2011). Factors associated with frequent utilization of crisis substance use detoxification services. *J Addict Dis, 30*(2), 116-122. doi: 936254277 [pii]10.1080/10550887.2011.554776
Carroll, C. P., Triplett, P. T., & Mondimore, F. M. (2009). The Intensive Treatment Unit: A brief inpatient detoxification facility demonstrating good postdetoxification treatment entry. *J Subst Abuse Treat, 37*(2), 111-119. doi: S0740-5472(08)00216-X [pii]10.1016/j.jsat.2008.11.003

Ford, L, & Zarate, P. (2010). Closing the gaps: The impact of inpatient detoxification and continuity of care on client outcomes. *Journal of Psychoactive Drugs, SARC Supplement 6, September*, 303-314.
Kertesz, S. G., Horton, N. J., Friedmann, P. D., Saitz, R., & Samet, J. H. (2003). Slowing the revolving door: stabilization programs reduce homeless persons' substance use after detoxification. *J Subst Abuse Treat, 24*(3), 197-207.
Lee, M. T., Horgan, C. M., Garnick, D. W., Acevedo, A., Panas, L., Ritter, G. A., . . . Reynolds, M. (2014). A performance measure for continuity of care after detoxification: relationship with outcomes. *J Subst Abuse Treat, 47*(2), 130-139. doi: 10.1016/j.jsat.2014.04.002
Mark, T. L., Dilonardo, J. D., Chalk, M., & Coffey, R. (2003). Factors associated with the receipt of treatment following detoxification. *J Subst Abuse Treat, 24*(4), 299-304. doi: S0740547203000394 [pii]
Mark, T. L., Vandivort-Warren, R., & Montejano, L. B. (2006). Factors affecting detoxification readmission: analysis of public sector data from three states. *J Subst Abuse Treat, 31*(4), 439-445. doi: S0740-5472(06)00161-9 [pii]10.1016/j.jsat.2006.05.019
McCollister, K. E., & French, M. T. (2003). The relative contribution of outcome domains in the total economic benefit of addiction interventions: a review of first findings. *Addiction, 98*(12), 1647-1659.
McCollister, K. E., French, M. T., & Fang, H. (2010). The cost of crime to society: new crime-specific estimates for policy and program evaluation. *Drug Alcohol Depend, 108*(1-2), 98-109. doi: 10.1016/j.drugalcdep.2009.12.002
McCusker, J., Bigelow, C., Luippold, R., Zorn, M., & Lewis, B. F. (1995). Outcomes of a 21-day drug detoxification program: retention, transfer to further treatment, and HIV risk reduction. *Am J Drug Alcohol Abuse, 21*(1), 1-16.
McLellan, A. T., Weinstein, R. L., Shen, Q., Kendig, C., & Levine, M. (2005). Improving continuity of care in a public addiction treatment system with clinical case management. *Am J Addict, 14*(5), 426-440. doi: M165J02445LNQ613 [pii]10.1080/10550490500247099
Stein, B. D., Kogan, J. N., & Sorbero, M. (2009). Substance abuse detoxification and residential treatment among Medicaid-enrolled adults: rates and duration of subsequent treatment. *Drug Alcohol Depend, 104*(1-2), 100-106. doi: S0376-8716(09)00137-9 [pii]10.1016/j.drugalcdep.2009.04.008
Thakur, N. M., Hoff, R. A., Druss, B., & Catalanotto, J. (1998). Using recidivism rates as a quality indicator for substance abuse treatment programs. *Psychiatric services (Washington, D C), 49*(10), 1347-1350.

[Response Ends]

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

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

[Response Begins]
There is general agreement that continuity of care should occur within a short time after discharge from detox (American Society of Addiction Medicine, 2014). By reporting both a 7- and 14-day follow-up timeframe, the measure balances clinical best practice thinking while recognizing the system's capacity limitations that could cause longer than optimal follow-up timeframes.
[Response Ends]

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

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

[Response Begins]
The measure was tested using the Transformed Medicaid Statistical Information System (T-MSIS) data and Medicare Part A, B, C and D claims data from calendar year 2018. Nine states representing a cross section of US regions (1 Northeast, 2 Midwest, 4 South, 2 West), were used for measure testing: State 1 South, State 2 South, State 3 Midwest, State 4 Midwest, State 5 West, State 6 West, State 7 Northeast, State 8 South, and State 9 South. These states have been blinded to protect confidentiality.

For the 9 states used in measure testing, there were 52,351 beneficiaries who had 75,830 discharges from a medically managed withdrawal program (denominator) during the measure period (January 1, 2018 through December 15, 2018). Beneficiaries who received follow-up within 14 days or 7 days of discharge are compliant for this measure.
The number of beneficiaries, number of medically managed withdrawal episodes (denominator), the 14-day measure performance rate and 7-day measure performance rates overall and for each state are shown below. State-level 14-day measure performance rates ranged from 24.9% to 57.1%, and 7-day measure performance rates ranged from 19.9% to 52.8%.
Overall (Nine States)
Number of beneficiaries: 52,351
Number of medically managed withdrawal episodes (denominator): 75,830
14-day Performance Rate
Overall (Nine States)
Weighted* mean performance rate: 47.30%
Unweighted mean performance rate: 41.01%
Standard deviation: 11.78%
Minimum: 24.91%
10th percentile: 26.27%
25th percentile: 34.50%
50th percentile: 41.00%
75th percentile: 52.64%
90th percentile: 54.21%
Maximum: 57.11%
IQR: 18.14
*Weighted average across states by denominator (number of medically managed withdrawal episodes)
7-Day Performance Rate
Overall (Nine States)
Weighted* mean performance rate: 42.32%
Unweighted mean performance rate: 35.70%
Standard deviation: 12.29%
Minimum: 19.96%
10th percentile: 21.46%
25th percentile: 27.85%
50th percentile: 33.86%
75th percentile: 47.94%
90th percentile: 50.27%
Maximum: 52.78%
IQR: 20.09
*Weighted average across states by denominator (number of medically managed withdrawal episodes)
The number of beneficiaries, number of medically managed withdrawal episodes (denominator), the 14-day and 7-day performance rates and their 95% confidence intervals (95% CI) are shown below.
State 1 South:
-Number of beneficiaries: 2,803
-Number of medically managed withdrawal episodes (denominator): 4,780
-14-day Performance rate: 34.50%. 95% CI: 33.15%, 35.85%
- 7-day Performance rate: 28.22%. 95% CI: 26.95%, 29.50%
State 2 South:
-Number of beneficiaries: 3,665
-Number of medically managed withdrawal episodes (denominator): 4,598
-14-day Performance rate: 57.11%. 95% CI: 55.68%, 58.54%
- 7-day Performance rate: 52.78%. 95% CI: 51.34%, 54.23%
State 3 Midwest:
-Number of beneficiaries: 11,271
-Number of medically managed withdrawal episodes (denominator): 15,765
-14-day Performance rate: 53.49%. 95% CI: 52.71%, 54.26%
- 7-day Performance rate: 49.64%. 95% CI: 48.85%, 50.42%
State 4 Midwest:
-Number of beneficiaries: 1,599

-Number of medically managed withdrawal episodes (denominator): 2,000
-14-day Performance rate: 34.65%. 95% CI: 32.56%, 36.74%
- 7-day Performance rate: 27.85%. 95% CI: 25.89%, 29.81%
State 5 West:
-Number of beneficiaries: 1,233
-Number of medically managed withdrawal episodes (denominator): 2,545
-14-day Performance rate: 24.91%. 95% CI: 23.23%, 26.59%
- 7-day Performance rate: 19.96%. 95% CI: 18.41%, 21.51%
State 6 West:
-Number of beneficiaries: 6,292
-Number of medically managed withdrawal episodes (denominator): 9,449
-14-day Performance rate: 41.00%. 95% CI: 40.01%, 41.99%
- 7-day Performance rate: 33.86%. 95% CI: 32.90%, 34.81%
State 7 Northeast:
-Number of beneficiaries: 20,596
-Number of medically managed withdrawal episodes (denominator): 30,157
-14-day Performance rate: 52.64%. 95% CI: 52.08%, 53.20%
- 7-day Performance rate: 47.94%. 95% CI: 47.38%, 48.50%
State 8 South:
-Number of beneficiaries: 3,223
-Number of medically managed withdrawal episodes (denominator): 4,558
-14-day Performance rate: 26.61%. 95% CI: 25.33%, 27.90%
- 7-day Performance rate: 21.83%. 95% CI: 20.63%, 23.03%
State 9 South:
-Number of beneficiaries: 1,669
-Number of medically managed withdrawal episodes (denominator): 1,978
-14-day Performance rate: 44.19%. 95% CI: 42.00%, 46.37%
- 7-day Performance rate: 39.18%. 95% CI: 37.03%, 41.33%
Data Sources: Transformed Medicaid Statistical Information System (T-MSIS) Analytic files (TAF) Research Identifiable Files (RIFs): demographic and eligibility (DE), other services (OT), inpatient (IP), long-term care (LT) and pharmacy (Rx). Medicare Advantage (MA) encounter files and Medicare Fee-For-Service files: master beneficiary summary file (MBSF), carrier, outpatient (OP), inpatient (IP), skilled nursing facility (SNF), MedPAR and prescription drug event (PDE).

[Response Ends]

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

[Response Begins]
Not applicable. Data have been included in Section 1b.2; these data represent continuity rates from 9 states included in testing using T-MSIS TAF RIF data from 2018.
[Response Ends]

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

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

[Response Begins]

During measure testing, the 14-day and 7-day measure performance rates were stratified by age group, race/ethnicity, dual eligibility for Medicare and Medicaid, gender, and Medicaid eligibility category. Disparities by these characteristics were similar for the 14-day and 7-day performance rates.

Measure performance for the 9 states, stratified by age group

Age 18–24 years

-Number of beneficiaries: 3,980
-Number of medically managed withdrawal episodes (denominator): 5,850
-14-day performance rate: 48.02%
-7-day performance rate: 43.21%

Age 25–44 years

-Number of beneficiaries: 31,008
-Number of medically managed withdrawal episodes (denominator): 45,528
-14-day performance rate: 49.24%
-7-day performance rate: 44.31%

Age 45–64 years

-Number of beneficiaries: 17,363
-Number of medically managed withdrawal episodes (denominator): 24,452
-14-day performance rate: 43.52%
-7-day performance rate: 38.40%

Measure performance rates for the 9 states, stratified by race/ethnicity

White non-Hispanic

-Number of beneficiaries: 33,731
-Number of medically managed withdrawal episodes (denominator): 49,499
-14-day performance rate: 48.91%
-7-day performance rate: 43.83%

Black non-Hispanic

-Number of beneficiaries: 9,242
-Number of medically managed withdrawal episodes (denominator): 12,485
-14-day performance rate: 48.27%
-7-day performance rate: 44.04%

Hispanic/Latino

-Number of beneficiaries: 5,256
-Number of medically managed withdrawal episodes (denominator): 7,829
-14-day performance rate: 41.17%
-7-day performance rate: 35.51%

Other or Unknown race and ethnicity:

-Number of beneficiaries: 4,122
-Number of medically managed withdrawal episodes (denominator): 6,017
-14-day performance rate: 40.02%
-7-day performance rate: 35.22%

Measure performance for the 9 states, stratified by gender

Male

-Number of beneficiaries: 20,290
-Number of medically managed withdrawal episodes (denominator): 28,236
-14-day performance rate: 48.58%
-7-day performance rate: 43.40%

Female

-Number of beneficiaries: 32,061
-Number of medically managed withdrawal episodes (denominator): 47,594
-14-day performance rate: 46.54%
-7-day performance rate: 41.68%

Measure performance for the 9 states, stratified by dual eligibility for Medicare and Medicaid

Medicaid only (Non-dual)

-Number of beneficiaries: 45,728
-Number of medically managed withdrawal episodes (denominator): 66,543
-14-day performance rate: 49.67%
-7-day performance rate: 44.70%

Dual eligible
-Number of beneficiaries: 6,623
-Number of medically managed withdrawal episodes (denominator): 9,287
-14-day performance rate: 30.38%
-7-day performance rate: 25.24%
Measure performance for the 9 states, stratified by Medicaid eligibility category.
Adult
-Number of beneficiaries: 38,534
-Number of medically managed withdrawal episodes (denominator): 56,287
-14-day performance rate: 50.66%
-7-day performance rate: 45.68%
Disabled
-Number of beneficiaries: 13,089
-Number of medically managed withdrawal episodes (denominator): 18,293
-14-day performance rate: 37.61%
-7-day performance rate: 32.73%
All other groups (CHIP, Child, Unknown)
-Number of beneficiaries: 850
-Number of medically managed withdrawal episodes (denominator): 1,250
-14-day performance rate: 38.00%
-7-day performance rate: 31.28%
Data Sources: Transformed Medicaid Statistical Information System (T-MSIS) Analytic files (TAF) Research Identifiable Files (RIFs): demographic and eligibility (DE), other services (OT), inpatient (IP), long-term care (LT) and pharmacy (Rx). Medicare Advantage (MA) encounter files and Medicare Fee-For-Service files: master beneficiary summary file (MBSF), carrier, outpatient (OP), inpatient (IP), skilled nursing facility (SNF), MedPAR and prescription drug event (PDE).

[Response Ends]

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

[Response Begins]
Not applicable. Performance data provided for **Question 1b.4**.
[Response Ends]

Criteria 2: Scientific Acceptability of Measure Properties

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

---

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

[Response Begins]
No
[Response Ends]

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

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

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

[Response Begins]
Not applicable
[Response Ends]

sp.01. Provide the measure title.

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

[Response Begins]
Continuity of Care After Medically Managed Withdrawal from Alcohol and/or Drugs
[Response Ends]

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

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

[Response Begins]
Percentage of discharges from a medically managed withdrawal episode for adult Medicaid Beneficiaries, age 18-64, that was followed by a treatment service for substance use disorder (including the prescription or receipt of a medication to treat a substance use disorder (pharmacotherapy) within 7 or 14 days after discharge). This measure is reported across all medically managed withdrawal settings.
[Response Ends]

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

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Surgery: General*

[Response Begins]
Behavioral Health: Alcohol, Substance Use/Abuse
Other (specify)
[Other (specify) Please Explain]
Home and Community-Based Services, including older adults, persons with a physical disability, persons with an intellectual or developmental disability (ID/DD), persons with an acquired brain injury (ABI), and persons with mental health or substance use disorders (MH/SUD).

[Response Ends]

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

[Response Begins]
Care Coordination
[Response Ends]

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

Select only those target populations which can be stratified in the reporting of the measure's result.
Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]
Adults (Age >= 18)
Populations at Risk: Dual eligible beneficiaries of Medicare and Medicaid
Populations at Risk: Populations at Risk
[Response Ends]

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

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.
Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

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

[Response Begins]
Population: Population
Population: Regional and State
[Response Ends]

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

Check ONLY the settings for which the measure is SPECIFIED and TESTED.
[Response Begins]
Inpatient/Hospital
Outpatient Services
[Response Ends]

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

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

[Response Begins]
https://www.medicaid.gov/state-resource-center/innovation-accelerator-program/iap-downloads/functional-areas/techspecsmanual-nqf-3312.pdf
[Response Ends]

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

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.
[Response Begins]
Available in attached Excel or csv file
[Response Ends]

Attachment: 3312_NQF-3312_Value_Sets.xlsx.xlsx

sp.12. State the numerator.

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

[Response Begins]
Discharges in the denominator who have an inpatient, intensive outpatient, partial hospitalization, outpatient visit, residential, or drug prescription or procedure within 7 or 14 days after discharge from an inpatient hospital, residential addiction program, or ambulatory medically managed withdrawal.
[Response Ends]

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

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

[Response Begins]
The numerator includes individuals with any of the following within 14 days after discharge from medically managed withdrawal:
• Pharmacotherapy on day of discharge through day 7 or 14.
• Outpatient, intensive outpatient, partial hospitalization, or residential treatment procedure with a diagnosis of SUD on the day after discharge through day 7 or 14.
• Outpatient, intensive outpatient, partial hospitalization, or residential treatment with standalone SUD procedure on the day after discharge through day 7 or 14.
• Inpatient admission with an SUD diagnosis or procedure code on day after discharge through day 7 or 14.
• Long-term care institutional claims with an SUD diagnosis on day after discharge through day 7 or 14.
If an overdose diagnosis code appears on the same outpatient or inpatient claim that is being viewed as follow-up, that claim does not qualify as follow-up.
SUD diagnoses are used to identify procedures connected to SUD diagnoses. SUD diagnoses are identified through ICD-10 codes (available in the attached value set: NQF 3312—Tab 3). Procedures are defined using a combination of Healthcare Common Procedure Coding System (HCPCS) codes, Uniform Billing (UB) Revenue Codes and ICD-10 procedure codes (available in the attached value sets: NQF 3312—Tabs 4–8).

Pharmacotherapy includes naltrexone (short or long acting), acamprosate, or disulfiram for alcohol dependence treatment and buprenorphine for opioid dependence treatment, as well HCPCS codes to identify procedures related to injecting drugs (e.g., long-acting injectable naltrexone) (available in the attached value sets: NQF 3312—Tabs 9–10). Code lists for this measure are in the attached value sets. States may need to adapt the list of codes to include state-specific codes.

[Response Ends]

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]
Adult Medicaid beneficiary discharges from medically managed withdrawal from January 1 to December 15 of the measurement year.
[Response Ends]

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

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

[Response Begins]
Measure data is reported annually (12 months). To account for the 14-day time period after discharge from medically managed withdrawal, the denominator period starts January 1 and ends December 15 of the measurement year.
Eligible population meets the following conditions:
• Medicaid beneficiaries aged 18-64 with at least one discharge from medically managed withdrawal during the year January 1 to December 15.
• Enrolled in Medicaid during the month of discharge medically managed withdrawal and the following month.
The denominator is based on discharges, not individual beneficiaries. A beneficiary may have more than one qualifying medically managed withdrawal episode.
The location of medically managed withdrawal can include hospital inpatient, inpatient residential addiction, other stayover treatment, and ambulatory medically managed withdrawal. Medically managed withdrawal is identified using a combination of HCPCS codes, UB Revenue Codes, and ICD-10 procedure codes. A list of codes to identify medically managed withdrawal is posted in the value sets: Table NQF 3312—Tabs 1–2. States will likely need to modify the specifications to include their state-specific codes.

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]
Not applicable: the measure does not have denominator exclusions.
[Response Ends]

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

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

[Response Begins]
Not applicable: the measure does not have denominator exclusions.
[Response Ends]

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

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

[Response Begins]
Not applicable
[Response Ends]

sp.19. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.
[Response Begins]
No risk adjustment or risk stratification
[Response Ends]

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

Attachment: If available, please provide a sample report.
[Response Begins]
Rate/proportion
[Response Ends]

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

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score
[Response Begins]
Better quality = Higher score
[Response Ends]

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

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

[Response Begins]
The following step are used to identify the denominator, numerator, and calculation of the measure rate:
Step 1: Identify denominator
• Step 1A: Eligible population: Identify Medicaid beneficiaries ages 18–64 who have any medically managed withdrawal in inpatient hospital, residential addiction treatment program, or ambulatory medically managed withdrawal from January 1 to December 15 of the measurement year and are enrolled the month of medically managed withdrawal and the following month. Age is calculated as of January 1 of the measurement year.
• Step 1B: Among the Medicaid beneficiaries in Step 1A, identify all discharges from medically managed withdrawal using all inpatient, outpatient, and ambulatory claims files or tables that contain HCPCS or ICD-10 procedure codes and UB revenue codes (see NQF 3312 – Tab 1-2 for code lists). If more than one discharge from medically managed withdrawal in a year, treat each discharge from medically managed withdrawal as a separate episode, e.g., an inpatient hospital

medically managed withdrawal in January and an ambulatory medically managed withdrawal in July counts as two episodes.
o Step 1B.1: Multiple medically managed withdrawal claims that are up to 2 days apart are combined into a single episode. Sort the inpatient, outpatient, and ambulatory discharge from medically managed withdrawals by Beneficiary ID and service dates to ensure the discharges from these multiple data sources are in chronological order. Then combine claims that are up to 2 days apart while retaining all clinical fields from each episode.
• Step 1C: Identify appropriate location of medically managed withdrawal services: hospital inpatient, inpatient residential addiction, outpatient residential outpatient addiction, other stayover treatment, and ambulatory medically managed withdrawal. Use HCPCS medically managed withdrawal procedure codes to assign medically managed withdrawal location whenever possible; revenue center medically managed withdrawal will map to the hospital inpatient location when the revenue codes appear on an inpatient claim or table (see attached value set: NQF 3312 – Tab 2). Revenue center medically managed withdrawal will map to other stayover treatment when the revenue codes appear on a non-inpatient claim. If there is more than 1 medically managed withdrawal location when episodes are combined, assign the location using the first claim´s location. If there is a tie between a medically managed withdrawal episode being identified via revenue center codes and a more specific category using HCPCS on the same claim, the HCPCS location prevails.
Step 2: Identify numerator
• Step 2A: From the denominator in Step 1B, identify those discharges from medically managed withdrawal in any setting with a qualifying continuity service within 7 or 14 days after discharge.
o Step 2A.1: Identify SUD continuity services: Continuity services are assigned using clinical claims billing information (e.g., diagnosis, procedure, revenue codes; see attached value sets NQF 3312 – Tab 2-8). The measure includes all claims files or data tables that contain clinical fields (e.g., inpatient hospital, outpatient, other ambulatory, and long-term care). SUD diagnoses can be in any position – primary or secondary – for continuity services. Since multiple claims files or tables could each contain a continuity claim, this calls for creating continuity variables separately within each file type or table, sorting the files or tables by beneficiary ID and service dates, then putting them together in order to assign the set of variables that are "First" to occur relative to the medically managed withdrawal episode discharge date. Continuity services have to occur the day after discharge through day 7 or 14.
o Step 2A.2: Identify pharmacotherapy which may occur in multiple files or tables (see attached value sets: NQF 3312 – Tab 9-10). For example, one claims file or data source may contain injectables, another claims file or table data source may contain oral medications. Consequently, pharmacotherapy variables are created separately in each source, the data sources are then sorted by beneficiary ID and service dates, then multiple pharmacotherapy data sources are put together so they will be in chronological order to assign "First" variables. Pharmacotherapy services could be provided on the same day as the discharge from medically managed withdrawal through day 7 or 14.
o Step 2A.3: Co-occurring events: Emergency department visits, even with an SUD diagnosis, do not count as continuity. Also, other continuity services, e.g., an outpatient visit that occur on the same day as an emergency department visit with an SUD diagnosis do not count as continuity. If an overdose diagnosis code appears on the same claim as the continuity service, then the service does not count as continuity. If an inpatient continuity claim has an emergency department visit meaning that the beneficiary was admitted through the emergency department, it is allowed to remain a continuity service.
Step 3: Calculate rate
• Step 3A: Calculate the overall 7- or 14-day continuity rates by dividing the number of discharges with a qualifying continuity service (Step 2A) by the denominator (Step 1B).
• Step 3B: Calculate the rates separately for each medically managed withdrawal location by dividing the respective number of discharges by each location with a qualifying continuity service (Step 2A) by the denominator (Step 1C).

[Response Ends]

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

[Response Begins]
Not applicable; this measure does not use a sample.
[Response Ends]

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

[Response Begins]
Claims
[Response Ends]

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

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

[Response Begins]
We used the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAFs) Research Identifiable Files (RIFs) which contains beneficiary, service utilization, administrative claims, and expenditure data for the Medicaid population, including those covered through both fee-for-service (FFS) and managed care payers. The Medicaid T-MSIS TAF RIFs for calendar year 2018 were used to assess importance, reliability and validity: Demographic and Eligibility (DE) file; Inpatient (IP) file; Other Services (OT) file; Long-Term Care (LT) file; and Pharmacy (RX) file. For beneficiaries dually enrolled in Medicare [Medicare Advantage (MA) or Medicare Fee-For-Service (FFS)], in addition to the T-MSIS TAF RIFs, we also used Medicare Part A, B, C and D administrative claims from the Center for Medicare and Medicaid's Chronic Conditions Warehouse (CCW): MA encounter and Medicare FFS carrier, outpatient (OP), inpatient (IP), skilled nursing facility (SNF), MedPAR and prescription drug event (PDE) files.
Measure validity was assessed through NCQA Healthcare Effectiveness Data and Information Set (HEDIS) data from measure year 2018. HEDIS data are collected from Medicaid Health Management Organizations and Preferred Provider Organizations via the NCQA Interactive Data Submission System (IDSS) portal.

[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]
No data collection instrument provided
[Response Ends]

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

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:
Current Submission:
Updated testing information here.
Previous Submission:
Testing from the previous submission here.

[Response Begins]
Yes
[Response Ends]

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

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

Testing from the previous submission here.

[Response Begins]
Yes
[Response Ends]

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

[Response Begins]
No
[Response Ends]

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

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

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

[Response Begins]
No additional risk adjustment analysis included
[Response Ends]

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

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

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.
2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.
2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.
2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).
2b3. For outcome measures and other measures when indicated (e.g., resource use):
• an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
OR
• rationale/data support no risk adjustment/ stratification.
2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;
OR
there is evidence of overall less-than-optimal performance.
2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.
2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.
2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:
2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and
2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.
(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions
Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.
Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
Risk factors that influence outcomes should not be specified as exclusions.
With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

2021 Submission:
Updated testing information here.
2018 Submission:
Testing from the previous submission here.

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

[Response Begins]
Claims
[Response Ends]

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

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

[Response Begins]
2021 Submission
The Transformed Medicaid Statistical Information System (T-MSIS) contains beneficiary, service utilization, administrative claims, and expenditure data for the Medicaid population, including those covered through both fee-for-service (FFS) and managed care payers. The team used the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF) Research Identifiable Files (RIFs) for calendar year 2018[1],[2] which have beneficiary-level claims and enrollment data for the Medicaid population covered through fee-for-service (FFS) and managed care payers. The measurement period to identify denominator events (discharge from a medically managed withdrawal program) was January 1, 2018, through December 15, 2018, which allows for a 7- and 14-day follow-up after the discharge. Among the available T-MSIS files, the following were accessed to assess importance, reliability, and validity:

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

For beneficiaries dually enrolled in Medicare [Medicare Advantage (MA) or Medicare Fee-For-Service (FFS)], in addition to the T-MSIS TAF RIFs, the following Medicare enrollment data and Part A, B, C and D administrative claims were accessed to assess importance, reliability, and validity:

- Medicare Master Beneficiary Summary file (MBSF);
- Medicare Outpatient (OP) Encounter file;
- Medicare Inpatient (IP) Encounter file;
- Medicare Carrier Encounter file;
- Medicare skilled nursing facility (SNF) Encounter file;
- Medicare Inpatient (IP) FFS file;
- Medicare Outpatient (OP) FFS file;
- Medicare skilled nursing facility (SNF) FFS file;
- Medicare Carrier FFS file;
- Medicare MEDPAR file; and
- Medicare Part D prescription drug event (PDE) file.

For validity analysis, the team calculated the correlation between NQF 3312 and related measures for the behavioral health population in the Healthcare Effectiveness Data and Information Set (HEDIS). HEDIS data are collected from Medicaid Health Management Organizations and Preferred Provider Organizations via the NCQA Interactive Data Submission System (IDSS) portal. The team used HEDIS data from measure year 2018.
2016 Submission
Medicaid Analytic eXtract (MAX) 2013 and 2014 eligible (EL), inpatient (IP), other services (OT), long-term care (LT) and drug (RX) files. The other services file contains facility and individual provider services data. 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. The inpatient file only contains inpatient hospital, sterilization, abortion and religious non-medical health care institution claims.

We used the following MAX Medicaid files to identify adult Medicaid beneficiaries with discharges from detox (denominator) and the qualifying substance use treatment services and pharmacotherapy (numerator):
Person Summary (PS): Person-level file, including Medicaid eligibility and demographic information
Inpatient (IP): Claims-level file, including information on inpatient hospital stays
Long-Term Care (LT): Claims-level file, including information on long-term care institutional stays (nursing facilities, intermediate care facilities for individuals with intellectual disabilities, psychiatric hospitals, etc.)
Other Therapy (OT): Claims-level file, including information on use of "other" services, such as home- and community-based service use
Prescription Drug (RX): Information on drugs and other services provided by a pharmacy
[1] Centers for Medicare & Medicaid Services (2020). *2017–2018 Medicaid and CHIP data now available.* Centers for Medicare & Medicaid Services Research Data and Assistance Center. Accessed from https://www.resdac.org/cms-news/2017-2018-medicaid-and-chip-data-now-available.
[2] Centers for Medicare & Medicaid Services (2020) *Data file search.* Centers for Medicare & Medicaid Services Research Data and Assistance Center. Accessed from https://www.resdac.org/cms-data?tid_1%5B%5D=2.

[Response Ends]

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

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

[Response Begins]
2021 Submission
01-01-2018–12-31-2018

[Response Ends]

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

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.
Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.
Please do not select:

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

[Response Begins]
Population: Regional and State
[Response Ends]

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

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

[Response Begins]
2021 Submission
Nine states representing a cross section of US regions (1 Northeast, 2 Midwest, 4 South, 2 West) were used for measure testing: State 1 South, State 2 South, State 3 Midwest, State 4 Midwest, State 5 West, State 6 West, State 7 Northeast, State 8 South, and State 9 South. These states have been blinded to protect confidentiality.
2016 Submission
We included 14 states in measure testing: State A, State B, State C, State 3, State 4, State 4, State D, State E, State F, State 7, State G, State 8, State H, State I, and State J.

[Response Ends]

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

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

[Response Begins]
2021 Submission
Included in the testing and analyses were 52,351 beneficiaries with at least one detoxification episode during the year. Table 1 shows the overall number and percent distribution by Medicaid beneficiary category, age, gender, and race/ethnicity for all states.

Table 1. Number and Percentage of Medicaid Beneficiaries in the Eligible Population for NQF 3312, and Number of Medically Managed Withdrawal Episodes Overall and by Key Demographic Characteristics

*	**Overall (9 States)**	*
*	***Number of Beneficiaries***	***Percentage of Beneficiaries***
Overall	52,351	*
Dually eligible for Medicare	*	*
Medicaid only (Non-Dual)	45,728	87.3
Dual-eligible	6,623	12.7
Age, years	*	*
18-24	3,980	7.6
25-44	31,008	59.2
45-64	17,363	33.2
Gender	*	*
Male	20,290	38.8
Female	32,061	61.2
Race and ethnicity	*	*
Black	9,242	17.7
Hispanic	5,256	10.0
Unknown/Other Race	4,122	7.9
White	33,731	64.4
Medicaid Bene category	*	*
Adults	38,534	73.6
CHIP	<11	NA
Children	232	0.7
Disabled	13,089	25.0
Unknown BENE CAT	614	1.2

* Cell intentionally left empty
Data Sources: Transformed Medicaid Statistical Information System (T-MSIS) Analytic files (TAF) Research Identifiable Files (RIFs): demographic and eligibility (DE), other services (OT), inpatient (IP), long-term care (LT) and pharmacy (Rx). Medicare Advantage (MA) encounter files and Medicare Fee-For-Service files: master beneficiary summary file (MBSF), carrier, outpatient (OP), inpatient (IP), skilled nursing facility (SNF), MedPAR and prescription drug event (PDE).
2016 Submission

Table 1 Testing population characteristics: Medicaid beneficiary category, age, gender, and race/ethnicity by states

Beneficiary Characteristics	Total	*	State A	*	State B	*	State C	*	State 3	*	State 4	*	State D	*	State E	*
*	Number of Benes with at least one Detoxification Episode	Distribution (%)	Number of Benes with at least one Detoxification Episode	Distribution (%)	Number of Benes with at least one Detoxification Episode	Distribution (%)	Number of Benes with at least one Detoxification Episode	Distribution (%)	Number of Benes with at least one Detoxification Episode	Distribution (%)	Number of Benes with at least one Detoxification Episode	Distribution (%)	Number of Benes with at least one Detoxification Episode	Distribution (%)	Number of Benes with at least one Detoxification Episode	Distribution (%)
TOTAL	47,313	100.0	2,028	100.0	1,255	100.0	336	100.0	3,315	100.0	798	100.0	618	100.0	4,734	100.0
Medicaid beneficiary category	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Aged	678	1.4	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.5	NR	0.0
Blind-disabled	13,832	29.2	267	13.2	679	54.1	53	15.8	1,027	31.0	540	67.7	348	56.3	836	17.7
Adult	31,886	67.4	1,758	86.7	569	45.3	283	84.2	2,234	67.4	248	31.1	262	42.4	3,854	81.4
Child	734	1.6	NR	0.1	NR	0.6	NR	0.0	15	0.5	NR	1.1	NR	0.8	38	0.8
Unknown	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0
Age	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
18-24	5,163	10.9	114	5.6	96	7.6	12	3.6	335	10.1	60	7.5	63	10.2	439	9.3
25-44	24,417	51.6	933	46.0	634	50.5	136	40.5	1,727	52.1	429	53.8	341	55.2	2,555	54.0
45-64	17,550	37.1	981	48.4	525	41.8	188	56.0	1,215	36.7	308	38.6	214	34.6	1,734	36.6
Gender	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Male	28,547	60.3	1,381	68.1	553	44.1	241	71.7	1,727	52.1	444	55.6	266	43.0	2,934	62.0
Female	18,583	39.3	647	31.9	702	55.9	95	28.3	1,550	46.8	353	44.2	352	57.0	1,794	37.9
Unknown	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0

Beneficiary Characteristics	Total	*	State A	*	State B	*	State C	*	State 3	*	State 4	*	State D	*	State E	*
Race/ethnicity	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
White	27,979	59.1	1,476	72.8	676	53.9	250	74.4	2,125	64.1	558	69.9	349	56.5	2,856	60.3
Black	9,449	20.0	241	11.9	268	21.4	24	7.1	862	26.0	201	25.2	215	34.8	958	20.2
American Indian/Alaskan Native	300	0.6	NR	0.2	NR	0.2	NR	2.4	29	0.9	NR	0.8	NR	0.2	NR	0.2
Asian	556	1.2	NR	0.4	NR	0.2	NR	0.3	NR	0.1	NR	0.0	NR	0.2	19	0.4
Hispanic/Latino	5,905	12.5	298	14.7	NR	0.3	NR	1.8	59	1.8	13	1.6	NR	0.6	331	7.0
Native Hawaiian/Pacific Islander	51	0.1	NR	0.0	NR	0.0	NR	0.0	NR	0.1	NR	0.0	NR	0.0	NR	0.0
Other Race/Ethnicity	201	0.4	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0
Unknown Race/Ethnicity	2,689	5.7	NR	0.0	303	24.1	47	14.0	197	5.9	19	2.4	48	7.8	554	11.7

* Cell intentionally left empty.

Beneficiary Characteristics	State F	*	State 7	*	State G	*	State 8	*	State H	*	State I	*	State J	*
*	Number of Benes with at least one Detoxification Episode	Percent Distribution	Number of Benes with at least one Detoxification Episode	Percent Distribution	Number of Benes with at least one Detoxification Episode	Percent Distribution	Number of Benes with at least one Detoxification Episode	Percent Distribution	Number of Benes with at least one Detoxification Episode	Percent Distribution	Number of Benes with at least one Detoxification Episode	Percent Distribution	Number of Benes with at least one Detoxification Episode	Percent Distribution
TOTAL	19,473	100.0	7,322	100.0	3,203	100.0	774	100.0	1,008	100.0	1,625	100.0	824	100.0
Medicaid beneficiary category	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Aged	NR	0.0	NR	0.0	NR	0.0	NR	0.0	662	65.7	NR	0.1	NR	0.0
Blind-disabled	3,419	17.6	4,013	54.8	1,051	32.8	517	66.8	84	8.3	830	51.1	168	20.4
Adult	15,819	81.2	2,905	39.7	2,067	64.5	246	31.8	233	23.1	754	46.4	654	79.4
Child	134	0.7	362	4.9	85	2.7	NR	1.2	29	2.9	39	2.4	NR	0.0
Unknown	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0
Age	*	*	*	*	*	*	*	*	*	*	*	*	*	*
18-24	1,402	7.2	1,343	18.3	542	16.9	91	11.8	263	26.1	342	21.0	61	7.4
25-44	9,255	47.5	4,120	56.3	1,986	62.0	382	49.4	592	58.7	839	51.6	488	59.2
45-64	8,724	44.8	1,817	24.8	675	21.1	299	38.6	153	15.2	444	27.3	273	33.1
Gender	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Male	14,459	74.3	3,361	45.9	1,109	34.6	359	46.4	545	54.1	604	37.2	564	68.4
Female	4,922	25.3	3,919	53.5	2,094	65.4	413	53.4	463	45.9	1,021	62.8	258	31.3
Unknown	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0	NR	0.0
Race/ethnicity	*	*	*	*	*	*	*	*	*	*	*	*	*	*
White	8,395	43.1	5,448	74.4	2,598	81.1	350	45.2	950	94.2	1,146	70.5	802	97.3
Black	4,905	25.2	1,278	17.5	292	9.1	78	10.1	15	1.5	92	5.7	20	2.4

Beneficiary Characteristics	State F	*	State 7	*	State G	*	State 8	*	State H	*	State I	*	State J	*
American Indian/Alaskan Native	84	0.4	13	0.2	NR	0.1	NR	0.1	NR	0.2	136	8.4	NR	0.0
Asian	495	2.5	12	0.2	NR	0.0	NR	0.5	NR	0.0	NR	0.6	NR	0.0
Hispanic/Latino	4,423	22.7	432	5.9	11	0.3	203	26.2	NR	0.5	116	7.1	NR	0.0
Native Hawaiian/Pacific Islander	36	0.2	NR	0.0	NR	0.1	NR	0.0	NR	0.0	11	0.7	NR	0.0
Other Race/ Ethnicity	NR	0.0	NR	0.0	201	6.3	NR	0.0	NR	0.0	NR	0.0	NR	0.0
Unknown Race/ Ethnicity	1,043	5.4	97	1.3	95	3.0	136	17.6	36	3.6	114	7.0	NR	0.0

* Cell intentionally left empty.
NR=Not reported; result is based on a cell size of 10 or fewer

[Response Ends]

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

[Response Begins]
Not applicable
[Response Ends]

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

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

[Response Begins]
2021 Submission
As described in section 1.6, we collected information on the following variables using data extracted from the T-MSIS Analytic files (TAF) Research Identifiable Files (RIFs) 2018 files: dual eligibility for Medicare and Medicaid, Medicaid eligibility category, age, gender and race/ethnicity. This measure is based on a process that should be carried out for all beneficiaries, so no adjustment for patient mix is necessary. We did collect information about these variables and assessed disparities in performance rate for each group.
2016 Submission

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

[Response Ends]

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

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

Choose one or both levels.
[Response Begins]
Accountable Entity Level (e.g., signal-to-noise analysis)
[Response Ends]

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

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

[Response Begins]
2021 Submission
We examined signal-to-noise reliability of NQF 3312 performance using the methodology described by John Adams.[1] This methodology uses a beta-binomial model to assess how well one can confidently distinguish the performance of one reporting entity from another. Conceptually, the beta-binomial model is the ratio of signal to noise. The signal is the proportion of the variability in measured performance that can be explained by real differences in performance across reporting entities (for NQF 3312, the reporting entities are states). The beta-binomial model is an appropriate tool when estimating the reliability of simple pass/fail rate measures, such as NQF 3312. Using this approach, reliability scores can range from 0.0 to 1.0, with a score of 0.0 implying that all variation is attributed to measurement error (noise), and a reliability of 1.0 implying that all variation is caused by a real difference in performance across reporting entities (signal). For NQF 3312, states are the reporting entity at which signal to noise reliability was assessed, as described in the formulas and explanations below.
The formula for signal-to-noise reliability is signal-to-noise reliability=$\sigma^2_{\text{state-to-state}}/(\sigma^2_{\text{state-to-state}} + \sigma^2_{\text{error}})$. Therefore, the team estimated two variances: 1) variance between states ($\sigma^2_{\text{state-to-state}}$), and 2) variance within states (σ^2_{error}).

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

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

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

$\hat{\mathbf{p}}$ is the observed rate for the state, and **n** is the state-specific denominator for the observed rate (in this context, the number of medically managed withdrawal episodes in the state).
Using Adams' methodology, the team estimated the reliability for each reporting entity (state), averaging reliability estimates across all reporting entities to produce a point estimate of signal-to-noise reliability (known as the *mean signal-to-noise reliability* value). Mean signal-to-noise reliability shows how well, on average, NQF 3312 differentiates between reporting entity performance for each metric.
Along with the point estimate of mean signal-to-noise reliability, the team also estimated:

1. **The mean, standard deviation, minimum and maximum signal-to-noise reliability for all states**. The standard deviation, minimum and maximum estimates of signal-to-noise reliability provide information about the stability of the reliability results. The narrower the range between the minimum and maximum estimates, the less the signal-to-noise reliability estimate will change due to idiosyncratic features of specific states.
2. **Key percentiles of the distribution (minimum, 25th, 50th, 75th, and maximum) for the state-level signal-to-noise reliability estimates**. Each state's reliability estimate is a ratio of signal to noise, as described above [that is, $\sigma^2_{\text{state-to-state}}/(\sigma^2_{\text{state-to-state}}+\sigma^2_{\text{error}})$]. Variability between states ($\sigma^2_{\text{state-to-state}}$) is the same for each state, while the specific state error (σ^2_{error}) varies. Reliability for each state is an ordinal measure of how well one can determine where a state lies in the distribution across states, with higher estimates indicating better reliability.

2016 Submission
We estimated SNR reliability for the SUD-5 measure using a beta-binomial model, which is suitable for binary pass/fail rate measures (Adams, 2009). For SUD-5, the pass/fail designation is defined as having or not having an eligible follow-up visit within a specified time frame (7 days and 14 days) after eligible discharge from a detoxification episode (an inpatient hospital, residential addiction program, or ambulatory detoxification). The beta-binomial model assumes the entity SNR score is a binomial random variable conditional on the entity's true value, which comes from the beta distribution (ranging from 0 t o1). We calculated SNR reliability in three steps (Adams, 2009, 2014):
First, we calculated state specific SUD-5 measure variance ("noise") as a function of the measure passing rate at the state level, [2] and the sample size, *n*:
(1);
Second, we used version 2.2 of the BETABIN SAS macro written by Wakeling (n/d) to fit the beta-binomial model to the SUD-5 dataset (Wakeling, n/d) . The macro produced the estimated average pass rate across all providers, as well as the Alpha () and Beta () parameters that describe the shape of the fitted beta-binomial distribution. We calculated the "signal" (between-state variation of the SUD-5 measure) using these parameters, as follows:
(2);
Third, we calculated the SNR reliability as the ratio of the between-level variance and the total variance (i.e., the sum of the between-level and within-level variances) of the SUD-5 measure rate:
; (3);
We calculated reliability of the SUD-5 measure using two alternative definitions of continuity treatment services after detoxification discharges set at 7 and 14 days as stated in the specifications.
Temporal consistency. We assessed the temporal consistency (also referred to as temporal stability) of the SUD-5 measure by examining the strength of association between the state-level measure results in four quarters of the 2014 measurement year. Specifically, we first aggregated the SUD-5 measure result for each state within each quarter in 2014, and then calculated Spearman's rank-order correlation coefficient (ranging from -1 to +1) between the state-level measure results in consecutive quarters (i.e. 2014 Q1 vs Q2, Q2 vs Q3, and Q3 vs Q4). High positive value indicates a strong tendency for the paired measure ranks to move together, whereas a negative value indicates that the paired measure ranks move in opposite directions.
Adams, JL (2009). The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation.
Beneficiaries who had an eligible follow-up visit within a specified timeframe/Eligible beneficiaries discharged from detox

[Response Ends]

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

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

[Response Begins]
2021 Submission
The signal-to-noise analysis was conducted for the 9 states overall and separately for each states for both the 14-day and 7-day continuity measures. NQF 3312 was highly reliable in distinguishing performance between states.
The average signal-to-noise reliability was 0.995 for the 14-day continuity rate, ranging from 0.990 to 0.999; for the 7-day continuity rate, the average signal-to-noise reliability was 0.996, and ranged from 0.991 to 0.999 (Tables 2a–2c).

Table 2a. Distribution of State Estimates of Signal-to-Noise Reliability for NQF 3312 14-Day Continuity of Care after Medically Managed Withdrawal from Alcohol and/or Drugs

Number of States	Mean	Standard Deviation	Minimum	25th percentile	50th percentile	75th percentile	Maximum
9	0.995	0.003	0.990	0.994	0.996	0.998	0.999

Table 2b. Distribution of State Estimates of Signal-to-Noise Reliability for NQF 3312 7-Day Continuity of Care after Medically Managed Withdrawal from Alcohol and/or Drugs

Number of States	Mean	Standard Deviation	Minimum	25th percentile	50th percentile	75th percentile	Maximum
9	0.996	0.003	0.991	0.995	0.997	0.998	0.999

Table 2c. State Estimates of Signal-to-Noise Reliability for NQF 3312 14-Day and 7-Day Continuity of Care after Medically Managed Withdrawal from Alcohol or Drugs

*	*	14-Day Continuity	*	*	*	7-Day Continuity	*
State Number	Denominator	Numerator	Rate	Signal-to-Noise Reliability	Numerator	Rate	Signal-to-Noise Reliability
State 1 South	4780	1649	34.50%	0.996	1349	28.22%	0.997
State 2 South	4598	2626	57.11%	0.996	2427	52.78%	0.996
State 3 Midwest	15765	8432	53.49%	0.999	7825	49.64%	0.999
State 4 Midwest	2000	693	34.65%	0.991	557	27.85%	0.992
State 5 West	2545	634	24.91%	0.994	508	19.96%	0.995
State 6 West	9449	3874	41.00%	0.998	3199	33.86%	0.998
State 7 Northeast	30157	15875	52.64%	0.999	14457	47.94%	0.999
State 8 South	4558	1213	26.61%	0.996	995	21.83%	0.997
State 9 South	1978	874	44.19%	0.990	775	39.18%	0.991

* Cell intentionally left empty.

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

2016 Submission

Signal-to-noise reliability. Table 2 summarizes the mean and range of the SNR statistic for SUD-5, which was computed separately for each of the 14 states in the sample by each of the two definitions of continuity threshold (7-day and 14-days; Table 3). Note that the threshold definition
only affects the SUD-5 measure numerator (eligible follow-up visit within a specified continuity timeframe) but does not affect the measure denominator (eligible discharge from a detoxification episode). Generally, we observed smaller numerator counts using 7-day continuity threshold compare to the 14-day continuity threshold.[1]

The SUD-5 was highly reliable in distinguishing performance between States using both 7- and 14-days continuity threshold, with the average reliability score of 0.98 across states and a range from 0.98 to 0.99.

Note that high reliability is not indicative of high quality of healthcare, but rather indicates that the SUD-5 measure can be used to distinguish between entities with respect to healthcare quality. The high reliability for the measure at the state level is likely driven by the adequate sample sizes and low "noise" variance within the States. The figure below demonstrates the relationship between the number of discharges from detox at the state level (the SUD-5 measure denominator) and the resulting SNR statistic.

Table 2. Signal-to-noise reliability for the SUD-5 measure (n=14 states)

SUD-5 Continuity Threshold	Average reliability score	Range of reliability scores
7-day	0.99	(0.99-0.99)
14-day	0.99	(0.98-0.99)

Alt text: This table compares the average reliability scores and range of reliability scores for both SUD-5 Continuity Thresholds, the 7- and 14-day thresholds. Both reliability scores are 0.99; the range of reliability score for 7-day is 0.99-0.99 while the range for 14-day is 0.98-0.99.

Notes: Data from 14 states were included in the analysis. Based on analysis of 2014 (2013 for State 8 and State I) MAX PS, IP, LT, OT, and RX files. For both 7- and 14-day continuity thresholds the signal-to-noise coefficients for State B, State H, State I, State A, State G, State D, State E, State 7 and State F were truncated to 0.99 rather than rounded to 1.00 to reflect the uncertainty in the estimates.

Table 3. SUD-5 Measure rate and signal-to-noise reliability, by State

State abbreviation	# of eligible discharges from detox	7-Day Continuity Threshold	*	*	14-Day Continuity Threshold	*	*
*	*	# of discharges with continuity treatment	Mean SUD-5 rate	Signal-to-noise reliability	# of discharges with continuity treatment	Mean SUD-5 rate	Signal-to-noise reliability
State C	379	78	0.21	0.99	107	0.29	0.98
State D	884	239	0.27	0.99	303	0.34	0.99
State 8	929	304	0.33	0.99	367	0.40	0.99
State J	985	380	0.39	0.99	448	0.46	0.99
State 4	997	289	0.29	0.99	365	0.37	0.99
State H	1,160	944	0.81	0.99	979	0.85	0.99
State B	1,571	287	0.18	0.99	409	0.26	0.99
State I	2,058	456	0.22	0.99	603	0.29	0.99
State A	2,799	858	0.31	0.99	1,162	0.42	0.99
State 3	3,760	2,368	0.63	0.99	2,537	0.68	0.99
State G	3,911	1,288	0.33	0.99	1,542	0.39	0.99
State E	6,068	959	0.16	0.99	1,405	0.23	0.99
State 7	9,474	4,882	0.52	0.99	5,544	0.59	0.99
State F	32,744	6,149	0.19	0.99	8,940	0.27	0.99

* Cell intentionally left blank

Notes: The signal-to-noise coefficients for VT, GA, WA, CT, MI, TN, NJ, PA and NY were truncated to 0.99 rather than rounded to 1.00 to reflect the uncertainty in the estimates.

Overall, using the 7-day continuity threshold we observed marginally higher signal to-noise-reliability for 13 out of 14 states (except VT for which the reliability decreased by 0.03 percentage points). On average, the 7-day continuity threshold average reliability was 0.1 percentage point (PP) higher (with the change in reliability ranging from -0.4 PP +4.7 PP) compared to the 14-day threshold.

The small increase in the SUD-5 measure reliability using the 7-day continuity threshold can be explained by examining three key drivers of reliability: "signal," "noise" and denominator sample size. First, using the 7-day continuity threshold results in somewhat larger between-state variance or stronger "signal" compared to the 14-day threshold (4 and 10 PP difference)l. Secondly, with the 7-day continuity threshold we observed smaller within-state variances (or weaker "noise") for 11 out of 14 states. The noise variance was on average 9.9 PP weaker. Since denominator size remained the same for both definitions of continuity threshold, consistently stronger "signal" with generally weaker "noise" mostly resulted in higher reliability of the SUD-5 data with the 7-day continuity threshold.

Temporal consistency. Table 4 provides the measures of temporal consistency (Spearman rank correlation) across four quarters of the 2014 measurement year for the SUD-5 measure. Our results indicate very high (at or above 0.90) temporal stability of the SUD-5 measure over time.

Table 4. Temporal consistency of SUD-5 in the measurement year

NQF 3312 Continuity Threshold	Spearman Rank Correlations	*	*	*
*	Average across 4 quarters	Q1 vs. Q2	Q2 vs. Q3	Q3 vs. Q4
7 Day	0.93	0.94	0.94	0.92
14 Days	0.93	0.96	0.92	0.92

* Cell intentionally left blank

Notes: All correlation coefficients are statistically significant at $p<0.001$; each pairwise correlation included only those States that had data during both quarters analyzed.

[1] For both 7- and 14-day continuity thresholds the average number of eligible discharges from detox per State for the SUD-5 measure was 4,837 (ranging from 379 to 32,744). The average number of episodes with continuity treatment within 7 days after discharge per State was 1.392 (ranging from 78 to 6149). The average number of episodes with continuity treatment within 14 days after discharge per State was 1,765 (ranging from 107 to 8,940). These statistics are

based on all eligible denominator and numerator counts in the SUD-5 dataset including those with less than 11 observations in the denominator or numerator.

[Response Ends]

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

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

[Response Begins]
2021 Submission
NQF 3312 is rated high for scientific acceptability, based on reliability results. The signal-to-noise analyses showed that the reliability of NQF 3312 is excellent for both 14-day and 7-day continuity performance measures. Although high signal-to-noise reliability is not indicative of high-quality health care, it does indicate that the measure may be used to distinguish between states with respect to health care quality.
High reliability for NQF 3312 is likely supported by large enough sample sizes at the state level. The average number medically managed withdrawal episodes was 8,425 (ranging from 2,545 to 30,157), and the average number of beneficiaries receiving continuity of care within 14 days of discharge from a medically managed withdrawal program was 3,986 (ranging from 634 to 15,875).
2016 Submission
SUD-5 is rated high for scientific acceptability, based on reliability testing results. Specifically, the excellent SNR indicated that the SUD-5 measure can discern the underlying performance between states within high precision. High temporal consistency showed that the performance of state-level SUD-5 rates were consistent over time.

[Response Ends]

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

[Response Begins]
Empirical validity testing
[Response Ends]

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

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

[Response Begins]
2021 Submission
Convergent validity. Validity for NQF 3312 14-day and 7-day continuity was evaluated in 2016 and demonstrated strong convergent validity with SUD-5 (please note, *SUD-5* refers to both rates for the NQF 3312 measure. Here, we performed convergent validity of the 14-day rate with the 7-day rate; all other references to *SUD-5* throughout this submission refer to NQF 3312 as a whole). The team also chose to explore additional approaches by comparing NQF 3312 state-level results with an existing HEDIS health plan measure assessing follow-up and emergency department (ED) visits for SUD. We share our approach and those results below in the spirit of transparency.
We calculated Spearman correlation coefficients for construct validity using state-level data derived from the Medicaid T-MSIS, TAF, and RIFs as well as HEDIS data, aggregated at the state level. Due to the number of states included in the analyses, we calculated Spearman rather than Pearson correlations. Spearman correlations estimate the strength and direction of the monotonic association between two continuous variables; the magnitude of correlation ranges from -1.0 to +1.0. A value of 1.0 indicates a strong positive association (that is, an increase in values of the first variable will be associated with an increase in value of the second variable). A value of 0.0 indicates no association between variables. A value of -1.0 indicates a strong negative association (that is, an increase in values of the first variable will be associated with a decrease in values of the second variable). The significance of a Spearman correlation coefficient is evaluated by testing the hypothesis that an observed coefficient calculated for the sample is different from zero. The resulting p value indicates the probability of obtaining a difference at least as large as the one observed due to chance alone. P values of less than 0.05 imply that it is unlikely a non-zero coefficient was observed due to chance alone.

NQF 3312 represents the percentage of discharges from a medically managed withdrawal episode for adult Medicaid beneficiaries, aged 18 to 64, that were followed by a treatment service for substance use disorder, including the prescription or receipt of a medication to treat a substance use disorder (pharmacotherapy), within 7 or 14 days after discharge. The team used the following research questions to evaluate NQF 3312:

- Are the 7- and 14-day indicators for NQF 3312 correlated with one another?
- Are the results for NQF 3312 correlated with the HEDIS *Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence* (FUA)[1] measure?

The team hypothesized that two rates for NQF 3312 will have a large and statistically significant positive correlation; we also hypothesize that states performing well on NQF 3312 will perform well on the selected HEDIS measure, as they address similar care coordination processes and access to care for shared populations (therefore demonstrating a positive rho, or correlation coefficient).

2016 Submission

We conducted validity testing at the performance score level for both 14- and 7- day continuity.

Convergent validity. To assess the convergent validity of the SUD-5 measure, we examined the association between presence/absence of continuity of care (that is, the underlying construct of the measure) - defined as having a follow-up visit within 7 or 14 days after discharge from detox—and presence/absence of a subsequent overdose treatment or detox readmission). We hypothesized that there would be fewer overdoses or detox readmissions between days 15 and 90 (for the analysis of 14-day continuity) or between days and 8 and 90 (for the analysis of 7-day continuity) after the detox among beneficiaries with continuity of care compared to those without. We used inverse probability weighting and doubly robust regression (Imbens & Wooldridge, 2009) to examine the association and controlled for potential confounders, including age group, gender, race/ethnicity, focus of detox, location of detox, blind/disabled status, the use of pharmacotherapy for SUD in the 30 days before the detox treatment of interest, and the use of behavioral and/or physical health services (in an inpatient, outpatient, or emergency department setting) in the 30 days before the detox treatment of interest.

[1] The HEDIS FUA measure assesses the proportion of ED visits for alcohol or other drug abuse or dependence that had a follow-up visit.

[Response Ends]

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

Examples may include correlations or t-test results.

[Response Begins]

2021 Submission

As expected, the NQF 3312 7-day and 14-day follow-up had a strong positive correlation (r=0.98, p<0.001).

Of the 9 states used for measure testing, only 6 states had HEDIS FUA Medicaid data available. The NQF 3312 14-day follow-up had a weak positive correlation with FUA (7-day follow-up: r=0.31, p=0.54; 30-day follow-up: r=0.14, p=0.78) (Table 3a). Similarly, NQF 3312 7-day follow-up had a weak positive correlation with FUA (7-day follow-up: r=0.37, p=0.47; 30-day follow-up: r=0.26, p=0.62). (Table 3b).

Table 3a. Spearman Correlation between NQF 3312 14-day follow-up and HEDIS FUA 7-Day and 30-Day Follow-up

Spearman Correlation with NQF 3312 Rate of 14-day follow-up	Number of States	Spearman Correlation	Correlation p-value
HEDIS FUA 7 Day: Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence	6	0.314	0.54
HEDIS FUA 30 Day: Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence	6	0.143	0.78

Table 3b. Spearman Correlation between NQF 3312 7-day follow-up and HEDIS FUA 7-Day and 30-Day Follow-up

Spearman Correlation with NQF 3312 Rate of 7-day follow-up	Number of States	Spearman Correlation	Correlation p-value
Measure	*	*	*

Spearman Correlation with NQF 3312 Rate of 7-day follow-up	Number of States	Spearman Correlation	Correlation p-value
HEDIS FUA 7 Day: Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence	6	0.371	0.47
HEDIS FUA 30 Day: Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence	6	0.257	0.62

* Cell intentionally left blank

2016 Submission

Convergent validity. The odds of subsequent overdose treatment or readmissions between days 15–90 among those with continuity of care within 14 days were 0.917 (with 95% confidence interval (0.863, 0.976)) as much as those of individuals without continuity of care, translating into an absolute risk reduction of 1.4 percent and a number needed to treat (NNT) of 71, which was statistically significant ($p < 0.01$). Similar results hold when looking into the outcome with a 7-day threshold. The odds of subsequent overdose treatment or readmissions between days 8–90 among those with continuity of care were 0.834 (with 95% confidence interval (0.783, 0.889)) as much as those of individuals without continuity of care, translating into an absolute risk reduction of 2.1 percent and an NNT of 48, which was statistically significant ($p < 0.01$).

[Response Ends]

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

[Response Begins]

2021 Submission

Prior analyses of convergent validity demonstrate NQF 3312 (known as *SUD-5* in other sections) to have high validity. NQF 3312 is also rated high for scientific acceptability, based on validity results. The convergent validity of NQF 3313 was excellent, with a lower odds (e.g., 8.3% lower for those with continuity of care within 14 days) of readmission to detox or overdose treatment among detox episodes with continuity. It is unsurprising that the additional exploratory convergent validity analysis of NQF 3312 14-day and 7-day follow-up with HEDIS FUA were not statistically significant at $p<0.05$ given the small sample size of N=6 in this analysis.

2016 Submission

SUD-5 is also rated high for scientific acceptability, based on validity results. The convergent validity of SUD-5 was excellent, with a lower odds (e.g. 8.3% lower for those with continuity of care within 14 days) of readmission to detox or overdose treatment among detox episodes with continuity.

[Response Ends]

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

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

[Response Begins]

2021 Submission

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

To determine if this difference is statistically significant, the team used an independent sample t-test of the performance difference between two randomly selected states: one with performance at or below the 25th percentile the other with performance at or above 75th percentile. This method takes into account the sample size, difference in performance rates and the variance of the performance rates. The test statistic is then compared against a normal distribution. If the p value of the test statistic is less than 0.05, then the two states' performance is significantly different from each other. Using this

method, we compared the performance rates of two randomly selected states, one state below the 25th percentile and another state above the 75th percentile of performance. We used these two states as examples of measures entities. However, the method can be used for comparison of any two measured entities.
Additionally, we compared measure performance for those dually eligible for both Medicare and Medicaid to those who are only eligible for Medicaid.

2016 Submission

We compared performance across state-level continuity rates to understand any variation in performance. We examined the distribution of the measure (for example, mean, median, minimum, 25th percentile, 75th percentile, and maximum) across states. In addition, we calculated the 95% confidence interval of the continuity rates for each state using a z-distribution for proportion. Then we compared each state's confidence interval to the overall measure rate that uses all beneficiaries across states. States measure rates significantly lower than the overall rate indicate an evidence of less-than-optimal performance, hence room for improvement.

[Response Ends]

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

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

[Response Begins]

2021 Submission

We found that performance rates across the 9 states covered a wide range with meaningful variation. Specifically, for NQF 3312 14-day follow-up, state-level performance rates ranged from 24.91 percent to 52.64 percent, with an average performance rate of 41.01 percent (Table 4a). For NQF 3312 7-day follow-up, state-level performance rates ranged from 19.96 percent to 52.78 percent, with an average performance rate of 35.70 percent (Table 4b).
The results of the t-test for NQF 3312 comparing a randomly selected state above the 75th percentile and a randomly selected state below the 25th percentile was statistically significant at $p<0.001$ for both the 7- and 14-day follow-up rates. This indicates that the difference in performance rates between the two states likely due to a true difference in performance and unlikely due to chance.

Table 4a. Variation in Performance on NQF 3312 14-Day Continuity of Care After Medically Managed Withdrawal from Alcohol and/or Drugs

14-Day Rate	*	*	*	*	*	*	*	*
Number of States	Mean number of detoxification episodes	Mean	Min	25th percentile	50th percentile	75th percentile	Max	IQR
9	8425.56	41.01	24.91	26.27	34.50	41.00	52.64	54.21

* Cell intentionally left blank

Table 4b Variation in Performance on NQF 3312 7-Day Continuity of Care After Medically Managed Withdrawal from Alcohol and/or Drugs

7-Day Rate	*	*	*	*	*	*	*	*
Number of States	Mean number of detoxification episodes	Mean	Min	25th percentile	50th percentile	75th percentile	Max	IQR
9	8425.56	35.70	19.96	27.85	33.86	47.94	52.78	20.09

* Cell intentionally left blank

As shown in Table 4c, across the 9 states included in measure testing, the 14-day and 7-day measure performance rates were 19 percentage points lower among beneficiaries dually eligible for both Medicare and Medicaid (representing 12% of the medically managed withdrawal episodes) compared to those eligible for Medicaid only.

Table 4c. Variation in Performance on NQF 3312 by Dual Eligibility Status

Dual Eligibility Status	Number of beneficiaries	Number of medically managed withdrawal episodes	NQF 3312 14-day performance rate	NQF 3312 7-day performance rate
Dually eligible for Medicare & Medicaid	6,623	9,287	30.38	25.24
Medicaid only	45,728	66,543	49.67	44.70

2016 Submission

For both 14-day and 7-day continuity rates, we found a wide range with meaningful variation. The 7-day continuity rate ranges from 15.80 percent to 81.38 percent with a median of 29.82 percent, and a mean of 34.46 percent (Table 5). The 14-day continuity rate ranges from 23.15 percent to 84.40 percent with a median of 38.02 percent, and a mean of 41.52 percent.

Table 5. Distribution of the SUD-5 Measure Rate

*	Distribution of the Measure Rate (%)	*	*	*	*	*
*	Minimum	25th Percentile	50th Percentile	75th Percentile	Maximum	Mean
7-day continuity	15.80	20.97	29.82	37.17	81.38	34.46
14-day continuity	23.15	28.50	38.02	44.49	84.40	41.52

* Cell intentionally left empty

Note: Data from 14 states are included in the analyses.

For both 7-day and 14-day continuity, the rates are greater than 50 percent in State H, State 3, and State 7 (Table 6 and Table 7). State E had the smallest measure rate.

For the 14-day continuity measure rate, the z-test for proportion indicates that 6 states have a measure rate significantly[1] greater than the overall measure rate, 5 states have a measure rate significantly lower than the overall measure rate, and the remaining 3 states had measure rates which were indistinguishable from the overall measure rate (Table 6).

Table 6. State-level SUD5 Measure Rate (14-day Continuity)

State	Number of Detoxification Episodes	Number of Detoxification Episodes with Continuity	Percentage of Detoxification Episodes with Continuity	95% Confidence Interval
Total	67,719	24,711	**36.49%**	*
State A*	2,799	1,162	41.51%	(39.69, 43.34)
State B†	1,571	409	26.03%	(23.86, 28.20)
State C†	379	107	28.23%	(23.70, 32.76)
State 3*	3,760	2,537	67.47%	(65.98, 68.97)
State 4	997	365	36.61%	(33.62, 39.60)
State D	884	303	34.28%	(31.15, 37.40)
State E†	6,068	1,405	23.15%	(22.09, 24.22)
State F†	32,744	8,940	27.30%	(26.82, 27.79)
State 7*	9,474	5,544	58.52%	(57.53, 59.51)
State G*	3,911	1,542	39.43%	(37.90, 40.96)
State 8	929	367	39.50%	(36.36, 42.65)
State H*	1,160	979	84.40%	(82.31, 86.48)

* Cell intentionally left empty

State	Number of Detoxification Episodes	Number of Detoxification Episodes with Continuity	Percentage of Detoxification Episodes with Continuity	95% Confidence Interval
State I†	2,058	603	29.30%	(27.33, 31.27)
State J*	985	448	45.48%	(42.37, 48.59)

Source: Based on analysis of 2014 (2013 for State 8 and State I) MAX PS, IP, LT, OT, and RX files.
Note: * Significantly greater than the total measure rate at the .05 level. † Significantly less than the total measure rate at the .05 level.
For the 7-day continuity, the z-test for proportion indicates that 7 states have a measure rate significantly greater than the overall measure rate, 5 states have a measures rate significantly lower than the overall measure rate, and the remaining 2 states have measure rates which were indistinguishable from the overall measure rate (Table 7).
Table 7. State-level SUD5 Measure Rate (7-day Continuity)

State	Number of Detoxification Episodes	Number of Detoxification Episodes with Continuity	Percentage of Detoxification Episodes with Continuity	95% Confidence Interval
Total	67,719	19,481	**28.77%**	*
State A*	2,799	858	30.65%	(28.95, 32.36)
State B†	1,571	287	18.27%	(16.36, 20.18)
State C†	379	78	20.58%	(16.51, 24.65)
State 3*	3,760	2,368	62.98%	(61.44, 64.52)
State 4	997	289	28.99%	(26.17, 31.80)
State D	884	239	27.04%	(24.11, 29.96)
State E†	6,068	959	15.80%	(14.89, 16.72)
State F†	32,744	6,149	18.78%	(18.36, 19.20)
State 7*	9,474	4,882	51.53%	(50.52, 52.54)
State G*	3,911	1,288	32.93%	(31.46, 34.41)
State 8*	929	304	32.72%	(29.71, 35.74)
State H*	1,160	944	81.38%	(79.14, 83.62)
State I†	2,058	456	22.16%	(20.36, 23.95)
State J*	985	380	38.58%	(35.54, 41.62)

Source: Based on analysis of 2014 (2013 for State 8 and State I) MAX PS, IP, LT, OT, and RX files.
* Significantly greater than the total measure rate at the .05 level.
† Significantly less than the total measure rate at the .05 level.
* Cell intentionally left empty
[Response Ends]

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

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

[Response Begins]
2021 Submission
These findings suggest room for improvement in follow-up after a medically managed withdrawal episode in the states included in testing. For the 7-day follow-up rate, eight of the nine states included had a performance of less than 50

percent, while six of the nine states had a performance of less than 50 for the 14-day follow-up rate. None of the states had performance rates above 53 percent and 58 percent for the 7-day and 14-day follow-up rates, respectively. Beneficiaries dually eligible for both Medicare and Medicaid had performance rates that were 19 percentage points lower than beneficiaries eligible for Medicaid only, indicating room for improvement.

2016 Submission

SUD-5 is rated high for validity, based on statistically significant and meaningful differences. The measure results suggest variation in performance and room for improvement in continuity of care after detoxification. For 14-day and 7- day continuity measures, five states had a continuity rate significantly below the overall total rate (Table 6 and Table 7). For the 14-day measure, three states had performance not distinguishable from the average performance; and for the 7-day measure, two states had performance not distinguishable from the average.

It is important to note that interpretation of the results should be tempered by the fact that only 14 states are included in the total; the total continuity rate for the entire nation could be different. The total is also weighted more heavily toward larger states, and states differ in terms of which detox and continuity services Medicaid covers. In terms of room for improvement, even states that are not statistically different from or even above the overall total of 36.5 percent have room for improvement. Achieving continuity of treatment after detoxification should be a goal for all clients, and only two states reach a rate greater than 60 percent for 7-day or 14-day continuity

[Response Ends]

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

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

[Response Begins]

2021 Submission

The team used multiple methods, including Data Quality Atlas (DQA) to identify states with sufficient data quality for inclusion in these analyses. Of the fifteen states included in the prior analysis, four had sufficient data quality to be included in this submission. We then examined the T-MSIS TAF RIFs calendar year 2018 to identify states with minimal missing data for the key variables to identify the numerator and denominator for NQF 3312. We identified an additional five states for a total of nine states that varied in size and geography for inclusion in these analyses.

2016 Submission

We assessed the extent of missing data was using the MAX validation and anomaly tables. These tables are available online at:

- MAX validation tables: https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAX-Validation-Reports.html?DLSort=0&DLEntries=10&DLPage=1&DLSortDir=ascending.
- MAX anomaly tables: https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html.

[Response Ends]

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

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

[Response Begins]

2021 Submission

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

2016 Submission

SUD-5 is a claims-based measure that relies on National Drug Code (NDC) in the RX file and procedure and diagnosis codes in the IP, LT, and OT files. Missing data is not a concern for many of the MAX data elements used to construct the SUD-5 measure in the study states.

- The service ending dates in the IP, OT, and LT files are non-missing because claims are assigned to yearly files using ending date; as such, a claim must have a non-missing ending date to be included in the MAX data. Similarly, prescription fill dates in the RX files are non-missing because RX claims are assigned to the yearly RX file using prescription fill date. Service beginning dates are infrequently missing.
- We found NDC to be non-missing in RX files.
- The SUD-5 specification utilizes secondary (and beyond) procedure and diagnosis codes; however, in the validation and anomaly tables, missing information is documented only for the primary diagnosis code and "a" procedure code. The absence of secondary primary and procedure codes may reflect missing data or may reflect the beneficiary's true clinical.
- Among the study states, the primary diagnosis code is mostly non-missing in the IP and LT files (Table 8). Missingness of primary diagnosis code in the OT file and procedure code in the IP and OT files varies by study state. For example, the percent of OT claims with a primary diagnosis code ranged from 57.5 percent in State I to 98.8 percent in State H (Table 8). In most states, most claims had a procedure code in the OT file. Procedure code in the IP file had higher rates of missingness in each state than in the OT file. Missing procedure and diagnosis codes may result in mistakenly excluding beneficiaries from the denominator or numerator, increasing the risk of over- or under-estimating the measure rate.

In State E and F, we found that the states were using state-specific codes for methadone treatment claims, which would not be currently captured by the measure specifications. In addition, State F frequently uses state-specific procedure codes. In the measure submission form, we advise measure implementers to include the relevant state-specific codes in the measure specification and calculation. Accounting for state-specific codes will improve the accuracy of measures calculated by states.

Table 8. Percent of IP, LT, or OT file with primary diagnosis code, procedure code, or place of service

*	**Percent with primary diagnosis code**	*	*	**Percent with procedure code**	*	**Percent with place of service**
State	IP	LT	OT	IP	OT	OT
State A	100.0	100.0	88.8	58.4	91.3	92.3
State B	100.0	100.0	95.7	60.5	96.3	88.3
State C	100.0	94.4	89.1	65.9	100.0	87.7
State 3	100.0	100.0	80.3	68.3	99.7	99.9
State 4	100.0	100.0	83.9	31.8	99.1	79.2
State D	100.0	100.0	97.5	42.9	100.0	93.1
State E	100.0	100.0	97.4	69.2	96.7	90.4
State F	100.0	100.0	97.4	74.8	99.2	87.6
State 7	100.0	100.0	97.3	67.4	100.0	75.7
State G	0.0	100.0	58.9	0.0	100.0	100.0
State 8	100.0	98.9	65.5	66.5	83.0	67.2
State H	100.0	100.0	98.8	58.3	91.6	92.9
State I	100.0	100.0	57.5	61.3	99.8	88.9
State J	100.0	100.0	90.7	59.7	98.9	96.4

* Cell intentionally left empty

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

Note: Numbers are from 2013 for all study states except State 8; the most recent numbers available for State 8 are from 2012.

We used two additional variables to create the measure – UB-92 revenue codes and place of service from the OT file. The percent of OT claims with a valid place of service ranges from 75.7 percent in State 7 to 100 percent in State G (Table 8). To calculate the SUD-5 measure generally and for specific subgroups, we also use data elements from the MAX PS file, including race, sex, zip code, age (calculated using date of birth), information about prepaid plans, and eligibility information. Sex and date of birth are rarely missing (Table 9). Nearly all enrollees have a valid 5-digit zip code. Race, however, is missing for a substantial portion of enrollees in some states (for example, 43.8 percent of enrollees in State C), so examination of SUD-5 by race subgroup will exclude beneficiaries who are missing race data. Information about

prepaid plans are generally non-missing. Over 95 percent of MAX claims have corresponding Medicaid eligibility information (Table 10).

Table 9: Percent of Medicaid enrollees with missing date of birth, sex, or race

State	Year	Percent of Enrollees Missing Date of Birth	Percent of Enrollees with Missing Sex	Percent of Enrollees with Missing Race
State A	2012	0.0	0.0	0.0
State B	2013	0.0	0.0	11.1
State C	2013	0.0	0.0	43.8
State 3	2012	0.0	0.0	10.7
State 4	2013	0.0	0.0	6.1
State D	2012	0.0	0.0	4.1
State E	2012	0.0	0.0	28.0
State F	2013	1.3	1.0	7.7
State 7	2013	0.0	0.0	12.3
State G	2013	0.0	0.0	10.9
State 8	2012	0.0	0.0	60.5
State H	2013	0.0	0.0	26.2
State I	2013	0.0	0.0	31.2
State J	2013	0.0	0.0	1.5

Alt text: This table shows the percentage of Medicaid enrollees with missing biographical or demographic data for each state examined, including the percentage of enrollees with missing dates of birth, missing sex, and missing race. Across all states, rates of enrollees with missing race information was higher than enrollees with missing date of birth or sex information.

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

Table 10: Percent of claims missing corresponding Medicaid eligibility information

State	Year	% with Claims and Missing Medicaid Eligibility (Excludes S-CHIP Only)	IP: % Missing Eligibility and > $0 Paid (Excludes S-CHIP Only)	LT: % Missing Eligibility and > $0 Paid (Excludes S-CHIP Only)	OT: % Missing Eligibility and > $0 Paid (Excludes S-CHIP Only)
State A	2013	0.27	0.22	0.07	0.18
State B	2013	0.96	0.12	0.02	0.17
State C	2014	0.85	0.07	0.01	0.16
State 3	2013	0.17	0.14	0.04	0.01
State 4	2014	0.08	0.04	0.02	0.00
State D	2013	4.08	1.59	0.41	0.38
State E	2014	1.50	0.94	0.46	0.10
State F	2013	2.03	0.14	0.01	0.97
State 7	2014	0.35	0.21	0.02	0.07
State G	2013	0.11	0.54	0.02	0.04
State 8	2014	0.23	0.28	0.02	0.10
State H	2013	0.55	0.20	0.42	0.21
State I	2014	0.55	0.24	0.33	0.21
State J	2013	0.07	0.23	0.21	0.00
State A	2013	2.82	1.01	0.47	0.08
State B	2014	3.77	0.94	0.16	0.31

State	Year	% with Claims and Missing Medicaid Eligibility (Excludes S-CHIP Only)	IP: % Missing Eligibility and > $0 Paid (Excludes S-CHIP Only)	LT: % Missing Eligibility and > $0 Paid (Excludes S-CHIP Only)	OT: % Missing Eligibility and > $0 Paid (Excludes S-CHIP Only)
State C	2014	0.01	0.00	0.00	0.00
State 3	2013	0.39	0.00	0.00	0.03
State 4	2012	12.23	0.28	0.01	1.21
State D	2014	0.56	0.00	0.00	0.03
State E	2013	0.53	0.93	0.44	0.17
State F	2014	0.22	0.41	0.29	0.04
State 7	2013	1.45	0.17	0.01	0.16
State G	2013	3.60	0.14	0.01	0.19
State 8	2014	0.09	0.05	0.02	0.01

Source: MAX validation tables. Available at the following URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAX-Validation-Reports.html?DLSort=0&DLEntries=10&DLPage=1&DLSortDir=ascending.
Note: Missing information is available for all but one of the study states in 2013. The most recent available information for State 8 is for 2012. We have also provided 2014 information where available in the study states.

[Response Ends]

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

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

[Response Begins]
2021 Submission
Due to the state selection process used for this analysis, the likelihood of missing data is negligible and would not contribute to systematic bias.
2016 Submission
Given the relatively small amount of missing information, we don't believe there is systematic bias. In addition, states implementing the measure will likely have even less missing data than reported here because they will be able to account for their state-specific codes when constructing the measure.
References
Adams, J. L. (2009). The Reliability of Provider Profiling. A Tutorial. http://www.rand.org/pubs/technical_reports/TR653.html
Adams, J. L. (2014). Reliability-Testing Concepts. National Quality Forum presentation. Retrieved from www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=74717
Imbens, G. W., & Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature, American Economic Association, 47*(1), 5-86.
Wakeling, I. (n/d). SAS Macro for fitting Beta-Binomial models. Retrieved from http://www.qistats.co.uk/BetaBinomial.html

[Response Ends]

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

demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

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

[Response Begins]
No, there is only one set of specifications for this measure
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

Examples may include correlation, and/or rank order.

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

[Response Begins]
N/A or no exclusions
[Response Ends]

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

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

[Response Begins]
Not applicable—This measure has no exclusions.
[Response Ends]

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

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

[Response Begins]
Not applicable—This measure has no exclusions.
[Response Ends]

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

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

[Response Begins]
Not applicable—This measure has no exclusions.
[Response Ends]

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

[Response Begins]
No risk adjustment or stratification
[Response Ends]

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

[Response Begins]
[Response Ends]

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

[Response Begins]
Not applicable—This measure is not risk adjusted.
[Response Ends]

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

2b.27. Provide risk model discrimination statistics.

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

[Response Begins]
[Response Ends]

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

[Response Begins]
Not applicable—This measure is not risk adjusted.
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

Criteria 3: Feasibility

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

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

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

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

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.
[Response Begins]
ALL data elements are in defined fields in electronic claims
[Response Ends]

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

[Response Begins]
Not applicable.
[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]
Not applicable.
[Response Ends]

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

[Response Begins]
We identified 9 states ranging in size and geographical location across the U.S. We used a variety of tools, including Data Quality Atlas (DQA) to identify states with sufficient data quality for these analyses, focusing on the germane data files (e.g., inpatient and outpatient claims) and key fields within those data files such as discharge date, ICD diagnosis codes, HCPCS procedure codes and UBREV codes.
[Response Ends]

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

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

Attach the fee schedule here, if applicable.

[Response Begins]
Not applicable, no fees or licensing are currently required.
[Response Ends]

Criteria 4: Use and Usability

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

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

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

Name of program and sponsor

URL

Purpose

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

Level of measurement and setting

[Response Begins]
Quality Improvement with Benchmarking (external benchmarking to multiple organizations)
[Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Please Explain]
MEDICAID INNOVATION ACCELERATOR PROGRAM: In July 2014, the Centers for Medicare & Medicaid Services (CMS) launched the Medicaid Innovation Accelerator Program (IAP), a collaborative between the Center for Medicaid and CHIP Services (CMCS) and the Center for Medicare & Medicaid Innovation (CMMI). The goal of IAP was to improve the health and health care of Medicaid beneficiaries and to reduce costs by supporting states' ongoing payment and delivery system reforms. Medicaid IAP supported state Medicaid agencies to build capacity in key program and functional areas by offering targeted technical support, tool development, and cross-state learning opportunities. The goal of the Medicaid IAP Reducing Substance Use Disorders (SUD) area was to support states to introduce policy, program, and payment reforms to better identify individuals with SUD, expand coverage for effective treatment, enhance care and practices delivered to beneficiaries, and develop payment mechanisms for SUD services that will provide better outcomes. The IAP covered all 50 states. NQF 3312 is available for optional reporting by states.

Quality Improvement (Internal to the specific organization)
[Quality Improvement (Internal to the specific organization) Please Explain]
MEDICAID INNOVATION ACCELERATOR PROGRAM: In July 2014, the Centers for Medicare & Medicaid Services (CMS) launched the Medicaid Innovation Accelerator Program (IAP), a collaborative between the Center for Medicaid and CHIP Services (CMCS) and the Center for Medicare & Medicaid Innovation (CMMI). The goal of IAP was to improve the health and health care of Medicaid beneficiaries and to reduce costs by supporting states' ongoing payment and delivery system reforms. Medicaid IAP supported state Medicaid agencies to build capacity in key program and functional areas by offering targeted technical support, tool development, and cross-state learning opportunities. The goal of the Medicaid IAP Reducing Substance Use Disorders (SUD) area was to support states to introduce policy, program, and payment reforms to better identify individuals with SUD, expand coverage for effective treatment, enhance care and practices delivered to beneficiaries, and develop payment mechanisms for SUD services that will provide better outcomes. The IAP covered all 50 states. NQF 3312 is available for optional reporting by states.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]
Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

Quality Improvement (internal to the specific organization)
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
[Response Ends]

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

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

[Response Begins]
We provided assistance with measure implementation to interested stakeholders during a CMS-sponsored webinar on June 20, 2019.
[Response Ends]

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

[Response Begins]
During a webinar on June 20, 2019, webinar participants were provided with information on measure implementation, including steps to calculate measure performance. Participants were given the opportunity to have their questions answered by the steward and developer.
[Response Ends]

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

[Response Begins]
Not applicable: no feedback has been received on measure performance and implementation from measured entities.
[Response Ends]

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

[Response Begins]
We have not received feedback on measure performance and implementation from measure implementers.

[Response Ends]

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

[Response Begins]
We have not received feedback from other users.
[Response Ends]

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

[Response Begins]
Not applicable.
[Response Ends]

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

[Response Begins]
Adoption of this performance measure has the potential to improve the quality of care for Medicaid beneficiaries, who have a SUD, after they are discharged from a medically managed withdrawal program for alcohol and/or drugs. For the 4 states included in both the 2016 and 2022 submissions, 14-day continuity decreased by an average of 5.7 percentage points and 7-day continuity decreased by an average of 4.0 percentage points. While the exact reasons for this are unknown, it is possible that this is due to improvements in data quality between 2014 and 2018 testing data, clinical initiatives and policies at the state level, or other factors. For all 9 states included in this submission, the 14-day performance rate is 47.30 percent and the 7-day performance rate is 42.32 percent, showing improvement from previous testing results of 36.5 percent continuity within 14 days and 28.8 percent within 7 days. Despite this improvement, the performance rates indicate that there is still a gap in care. The process of selection and testing of this measure was guided by the priorities outlined in the SAMHSA National Behavioral Health Quality Framework (NBHQF). The NBHQF goals reflect an effort to harmonize and prioritize measures that reflect the core principles of SAMHSA, as well as support the CMS National Quality Strategy. Specifically, this measure will encourage detox facilities to monitor the rate at which patients have follow-up treatment services to achieve continuity of care, and to take steps to put interventions in place to improve the rate with which their patients receive additional services after leaving detox.
[Response Ends]

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

[Response Begins]
To date, there have been no unexpected findings identified during use of this measure.
[Response Ends]

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

[Response Begins]
To date, there have been no unexpected benefits identified during use of this measure.
[Response Ends]

Criteria 5: Related and Competing Measures

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

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

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

(Can search and select measures.)
[Response Begins]
2605: Follow-Up After Emergency Department Visit for Mental Illness or Alcohol and Other Drug Abuse or Dependence
0004: Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment
3453: Continuity of Care after Inpatient or Residential Treatment for Substance Use Disorder (SUD)
[Response Ends]

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

(Can search and select measures.)
[Response Begins]
[Response Ends]

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

[Response Begins]
- 7-day Follow-up after Withdrawal Management; American Society of Addiction Medicine
- Continuity of Care after Detoxification; Washington Circle
- Initiation after Outpatient/Intensive Outpatient; Washington Circle

[Response Ends]

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

[Response Begins]
No
[Response Ends]

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

[Response Begins]
Follow-up time period: NQF 2605 examines follow-up care 7 days and 30 days after discharge. Similar to NQF 3312, NQF 3453 examines continuity of care 7 and 14 days after discharge from inpatient or residential treatment for SUD. NQF 3312 focuses on continuity of care 7 days and 14 days after medically managed withdrawal. The 14-day follow-up time period aligns with NQF 0004, NQF 3453 and the non-NQF endorsed Continuity of Care After Detoxification measure developed by the Washington Circle, and reflects the input of some public commenters that adults should receive some type of care within two weeks of discharge from detoxification.
Place of Service: NQF 3312 differs from the other related measures by examining continuity of care after medically managed withdrawal by place of service. NQF 2605 focuses on emergency department visit follow-up while NQF 0004

includes initiation of treatment through an inpatient AOD admission, outpatient visit, intensive outpatient encounter or partial hospitalization, telehealth, or medication treatment. NQF 3453 assesses follow-up after discharges from inpatient or residential treatment for substance use disorder.
Diagnoses: NQF 2605 requires a primary diagnosis of alcohol and other drug dependence (AOD) for the follow-up service. NQF 3453 numerator for follow-up requires a prescription for, administered, or ordered a medication for SUD. The denominator for NQF 3453 requires a discharge from inpatient or residential treatment for SUD with a principal SUD diagnosis. NQF 3312 requires a primary or secondary diagnosis of AOD. We allow a primary or secondary AOD diagnosis to address potential inaccuracies in how AOD diagnoses are coded. For example, some providers may be concerned about the stigma associated with an AOD diagnosis and therefore code it as a secondary diagnosis. Also, for adults with co-occurring mental health and AOD disorders, the assignment of primary and secondary diagnoses can be challenging and sometimes arbitrary.
The differences in follow-up time period, location, and diagnoses between NQF 2605, NQF 3453 and NQF 3312 do not impact the measure's interpretability in which a higher rate is indicative of better quality. These measures rely on administrative data. The differences in measure specifications between 2605, 3453 and 3312 are minor and expected to have minimal impact on data collection burden.

[Response Ends]

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

Provide analyses when possible.

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