NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

To navigate the links in the worksheet: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0004

Measure Title: Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment

Measure Steward: National Committee for Quality Assurance

Brief Description of Measure: This measure assesses the degree to which the organization initiates and engages members identified with a need for alcohol and other drug (AOD) abuse and dependence services and the degree to which members initiate and continue treatment once the need has been identified. Two rates are reported:

- Initiation of AOD Treatment. The percentage of adolescent and adult members with a new episode of AOD abuse or dependence who initiate treatment through an inpatient AOD admission, outpatient visit, intensive outpatient encounter, partial hospitalization, telehealth or medication assisted treatment (MAT) within 14 days of the diagnosis.
- Engagement of AOD Treatment. The percentage of adolescent and adult members with a new episode of AOD abuse or dependence who initiated treatment and who had two or more additional AOD services or MAT within 34 days of the initiation visit.

Developer Rationale: This measure assesses the degree to which the organization initiates and engages members identified with a need for alcohol and other drug dependence (AOD) services. By providing data on access to AOD dependence treatment across care settings, this measure provides insight on how plans and their providers may need to target education efforts and assists patient in accessing care.

Numerator Statement: Initiation of AOD Treatment:

Initiation of treatment through an inpatient AOD admission, outpatient visit, intensive outpatient encounter or partial hospitalization, telehealth or medication treatment within 14 days of the diagnosis.

Engagement of AOD Treatment:

Initiation of AOD treatment and two or more additional AOD services or medication treatment within 34 days of the initiation visit.

Denominator Statement: Patients age 13 years of age and older as of December 31 of the measurement year who were diagnosed with a new episode of alcohol or other drug dependency (AOD) during the first 10 and ½ months of the measurement year (e.g., January 1-November 15).

Denominator Exclusions: Exclude members who had a claim/encounter with a diagnosis of AOD abuse or dependence (AOD Abuse and Dependence Value Set), AOD medication treatment (AOD Medication Treatment Value Set) or an alcohol or opioid dependency treatment medication dispensing event (Medication Treatment for Alcohol Abuse or Dependence Medications List; Medication Treatment for Opioid Abuse or Dependence Medications List) during the 60 days (2 months) before the IESD.

Exclude patients who use hospice services or elect to use a hospice benefit any time during the measurement year, regardless of when the services began.

Measure Type: Process

Data Source: Claims

Level of Analysis: Health Plan

Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Feb 08, 2016

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

The developer provides the following evidence for this measure:

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

Summary of prior review in 2012

- During the last maintenance review in 2012, the developer provided a summary of how treatment frequency and intensity of engagement is important for successful outcomes.
- The developer cited evidence from the 2009 VA/DoD Clinical Practice Guideline for the Management of Substance Use Disorders.
- The Committee agreed the measure is important because it seeks to increase access and quality of care. The Committee noted that evidence presented during the 2012 submission was limited and did not discuss the capacity of the health care system to identify and engage people in treatment.

However, the evidence did show that those who are engaged have lower addiction severity index (ASI) scores overtime.

Changes to evidence from last review

- The developer added 3 guideline reviews which generally support the importance of various modes of treatment for substance abuse including psychosocial, medication, and intensive outpatient treatments. The guidelines are substantial based on hundreds of observational and clinical trials. The descriptions by the developers call out many general quotes from those reviews, but they are quite broad and do not specifically connect the measure target (one or multiple follow-up visits within a year of detection across a broad spectrum of treatments) to downstream quality of care. Moreover, the presentation of the evidence would be more readable if all "recommendations" were identified by brief descriptions as well as numbers. As one concrete example: the ASAM guideline cited in the article notes that psychosocial therapy is important to couple with medication therapy, but the measure numerator seems sometimes equally inclusive of either together or separately (though the engagement critieria does sometimes require multi-modal therapy for numerator "credit").

☐ **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:**

Updates:

Logic Model

- Patient (>13 years old) diagnosed with substance use disorder (SUD)→ Patient initiates SUD treatment through an inpatient or partial hospitalization event, or an outpatient encounter including medication assisted treatment or telehealth visit→Patient completes two or more additional SUD treatment services within 34 days of the initiation visit→Patient successfully engages in treatment (intermediate step), which supports a pathway to treatment completion and SUD recovery or appropriate ongoing management (desired outcome).

Clinical Guidelines

- The developer provided three additional published guidelines (beyond the one originally reviewed VA/DoD 2009) to support the measure as well updates to the VA/DoD Guideline.
  1. American Psychiatric Association Practice Guideline for the Treatment of Patients with Substance Use Disorders: Second Edition (2006)
- Provides a body of evidence based on 1,063 studies that psychosocial care and pharmacological treatments reduce SUD morbidity and mortality.
  2. American Psychiatric Association Practice Guideline for the Pharmacological Treatment of Patients with Alcohol Use Disorder (2018)
- Evidence for this Clinical Practice Guideline is supported by a systematic review that includes 95 randomized clinical trials.
- The guideline provides recommendations for evidence-based treatments planning for individuals with alcohol use disorder. It includes recommendations for both nonpharmacological and pharmacological treatment, and overall supports treatment initiation and engagement.
  3. The ASAM National Practice Guideline for the Use of Medications in the Treatment of Addiction Involving Opioid Use (2015)
     - Thirty-four guidelines were included in the analysis.
     - The Guideline states that psychosocial treatment is recommended in conjunction with any pharmacological treatment of opioid use disorder. Psychosocial treatment is generally recommended for patients receiving opioid agonist treatment and should be offered with extended-release naltrexone.
     - The estimates of benefit and consistency across studies found that patients experience improved and reproducible outcomes after receiving psychosocial treatment.
  4. VA/DoD Clinical Practice Guideline for the Management of Substance Use Disorders (2015)

- This Guideline was based on 135 studies, which included both randomized control trials and systematic reviews.
- The Guideline provide varous recommendations:
    i. Referral to specialy substance use care for patient with substance use disorder
    ii. Offering pharmacological treatment (Acamprosate, Disulfiram, Naltrexone-oral or extended release , or Topiramate) for moderate to severe alcohol use disorder
    iii. Offering nonpharmacological therapy for alcohol use disorder (Behavioral Couples Therapy for alcohol use disorder, Cognitive Behavioral Therapy for substance use disorders, Community Reinforcement Approach, Motivational Enhancement Therapy, 12-Step Facilitation)
    iv. Offering Pharmacological therapy (Buprenorphine/naloxone or Methadone in an Opioid Treatment Program) for opioid use disorder
    v. Offering extended-release injectable naltrexone for select patients with opioid use disorder
- The authors also discuss improvements in secondary outcomes based on recommended treatment such as crime associated with substance use, social engagement and vocational productivity, transmittable diseases, and morbidity.

Exception to evidence
N/A

Questions for the Committee:

- Is the evidence presentation sufficiently critical, synthetic, and current for the purpose of supporting this measure?
- Is the evidence provided too general or broad to support this specific measure?
- Does the evidence appropriately distinguish between abuse and dependence?
- Does the measure use the right time frames for assessing initation and engagement, 14 and 34 days, respectively?

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → QQC presented, but not much on consistency, and not specific to measure (Box 4) → Quality of general evidence generally high, focus of evidence is the issue (Box 6) → Moderate

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

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

Maintenance measures – increased emphasis on gap and variation

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

- Data from 2016-2018 is provided and includes average performance, N = number of health plans, min, max, standard deviation and percentiles (and where applicable stratification by diagnosis).

  Average Scores for Initiation (%, n)

	2016	**2017**	**2018**
Medicaid	38.24, 161	40.87, 184	42.28, 186
Medicare	33.25, 397	33.42, 404	34.40, 408

	2016	2017	2018
Commercial	33.92, 405	33.70, 401	36.65, 384

Average Scores for Engagement (%, n)

	2016	2017	2018
Medicaid	10.31, 163	12.66, 186	13.55, 188
Medicare	3.14, 397	3.52, 404	4.21, 408
Commercial	12.65, 405	12.09, 402	13.40, 384

- The data shows slight improvement in performance rates from 2016-2018 across all insurance types with apparent room for additional improvement.
- Mean performance and distribution for the initiation of treatment (initiation indicator) was relatively similar among the Medicare, Medicaid and commercial products.
- For the Engagement indicator, performance was about 10 percentage points lower for Medicare than Medicaid and commercial products.
- 2018 data was stratified by diagnosis (alcohol, opioid, other drug).
    - For both the Initiation and Engagement indicators, slightly higher performance was among members with a diagnosis of opioid abuse and dependence than members with diagnoses of alcohol or other drug abuse and dependence.

Disparities

- This measure is not stratified to detect racial/ethnic/language disparities as the developer does not routinely collect such information.
- The developer stratified data by type of insurance (Medicare, Medicaid, private) and (absent explicity rationale) strongly recommends this practice when the data is available, but believes that the measure specifications should not require such adjustment.
- Findings from the CMS Office of Minority Health report, Racial and Ethnic Disparities in Health Care and Medicare Advantage (2016) indicated that there were disparites for initiation and engagement in treatment for Asian or Pacific Islander patients and Hispanic patients compared to White patients. Black patients were more likely than White patients to initiate treatment and as likely as White patients to engage in treatment.

Questions for the Committee:

- Does the gaps in care described warrant maintainence of this indicator as a national performance measure?
- Is the measure specific enough to isolate desirable levels/modalities of care for the demoninator of cases identified?
- Does the Committee agree that the developer's approach to disparities is appropriate?

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

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

1a. Evidence
Comments:
**This measure relates to outcome desired though it seems a bit double barrelled -- medication assisted treatment vs psychosocial treatment as sufficient indication of engagement in care. The research suggests that

it is both together that has the best results and the comingling of these double barrelled items might make data analyses (as noted there are over 900 codes) and direction to the providers challenging.
**No additional evidence
**Strong evidence based and logic model supports the process measure
**The evidence provided through the additional guidelines provides ample evidence of the effectiveness of the various treatment modalities that are part of the numerator, but less information and evidence about the importance of initiation and engagement in treatment.
**Meets evidence. No New studies I am aware of.

1b. Performance Gap

Comments:

**Performance data was included and overall shows slight improvements over time. disparities in care were not examined.
**Performance gap continues to be a significant issue. This measure needs to be incentivized by the developer; population is small; lack of initiation and engagement have significant effects on mortality, morbidity and costs.
**Yes, large gap and variation by state.
**Performance data provided suggests a slight increase in performance for both initiation and engagement, as well as performance differences between Medicaid, Medicare, and Commercial Insurance. Therefore, the measure does identify performance gaps as well as improvements. The developers discussed at length the reasoning behind not providing disparities data by sociodemographic categories, however this should be included in an analysis of performance.
**Yes there is a performance gap.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing

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

Reliability

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

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

Validity

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

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

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

Evaluators: NQF Staff

Review A

Evaluation of Reliability and Validity

Reliability

- The developer conducted performance measure score reliability testing using a beta-binomial model (Adams 2009) as the ratio of signal to noise.
- Initiation Indicator: Sampling and measurement was done at the plan level and included 408 Medicare health plans, 186 Medicaid health plans, and 384 commercial health plans.
- Engagement indicator: Included 408 Medicare health plans, 188 Medicaid health plans, and 384 commercial health plans.
- Reliability scores for both Initiation and Engagement were >= 0.94 indicating good reliability (i.e., high confidence that the measure discrimates between plan level variability and overall "noise" within plans).

Validity

- Performance score validity testing was performed.
- The developer explored whether Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment (IET) was positively correlated with the Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence measure (FUA) and whether the two indicators within the Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment measure were positively correlated with each other.
- Pearson Correlation Coefficients demonstrated a significant, moderate (0.08 to 0.6) and typically significant correlation for both relationships tested.
    - For the Medicaid population, only engagement in treatment had significant positive correlation with follow-up.
- The face validity of the measure was supported by a a technical expert panel (TEP) of 21 members.
- The difference between plan performance at the 25th and 75th percentile was statistically significant for both indicator rates across all product lines. This analysis, while mostly transparent, was not clear about how the distributions of the interquartile t-testing were created around those apparent subsets.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- Is there any concern about the strength of the correlations with other similar measures?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions of subjects, codes)?
- Does the Committee support the inclusion of pharmacotherapy and telehealth to improve face validity?
- Any concern about the specifications (i.e., diagnostic and treatment codes) as they are enumerated (i.e. is the list complete and correct) or employed (i.e., used) or regarding the inclusion of both abuse and dependence in the denominator?

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

Evaluation A: Scientific Acceptability

Measure Number: 0004

Measure Title: Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment

Type of measure:

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

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

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**

☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**

☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☐ **Facility** ☒ **Health Plan**

☐ **Population: Community, County or City** ☐ **Population: Regional and State**

☐ **Integrated Delivery System** ☐ **Other**

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

2. **Briefly summarize any concerns about the** measure specifications**.**
   - Changes were made to specifications since last endorsement.
     - Added dispensing of pharmacotherapy for treatment of alcohol and opioid abuse and dependence as appropriate initiation and engagement criteria.
     - Added "telehealth" to denominator and numerators.
     - Extended Engagement of AOD Treatment time frame from 30 days to 34 days.
   - The Standing Committee was previously concerned with the inclusion of both abuse and dependence diagnosis in the measure and the very broad use of codes (i.e., many across the spectrum of treatment modalities, the value set has more than 900 codes).

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐ **No**

   N/A

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

   Submission document: Testing attachment, section 2a2.2

   - Reliability of the measure score was tested using a beta-binomial model to calculate the signal to noise ratio at the health plan level.
   - Reliability scores range from 0.0 to 1.0.
   - A minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities.

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

   Submission document: Testing attachment, section 2a2.3

   Beta-binomial reliability (Plan-level)

Measure Indicator Rate	**Beta Binomial Reliability**		
	Commercial Product	Medicare Product	Medicaid Product
Initiation	0.97	0.99	0.99
Engagement	0.94	0.96	0.99

   - Reliability scores indicate good reliability in the sense the variability in the measure is principally tied to between rather than within plan differences

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

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

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

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

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

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

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

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

    - A reasonable method was used for relaibity testing and scores indicate good reliability for both Initiation and Engagement in all product lines.

- Previous testing included a breakdown of reliability for by age group (13-17 years and 18 years and older). In the previous testing, for the Initiation indicator in the commercial product line, 13-17 year group the realiability score were slightly lower. Breaking down reliability results by age group would provide more detailed reliability information.
- This new submission does not address the distinction between "abuse" and "dependence", even as the 2012 submission of this measure did say this would be addressed in the future with ICD-10 code use. Does this issue need to be directly addressed? This likely will be of relevance to the denominator of the measure, or to anticipated effects with the numerator (i.e., more severe addiction would increase the need for initiation and engagement).

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    - Exclusions
        - Members who had a claim/encounter with a diagnosis of AOD abuse or dependence, AOD medication treatment, or an alcohol or opioid dependency treatment medication dispensing event during the 60 days before the IESD (the initial episode start date).
        - Patient who use hospice services or a hospice benefit during the measurement year
    - Testing was not performed for those subjects exclusions.

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

    Submission document: Testing attachment, section 2b4.

    Numerator of the measure is complex, thus lacking specificity to a proper treatment-diagnosis pairing

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    Submission document: Testing attachment, section 2b5.
    N/A

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

    Submission document: Testing attachment, section 2b6.

    - Claims data is said to be complete regarding the fields of interest. NCQA audits the diagnostic and procedure code fields in question. Concerns might still exist regarding the completeness of entry from the clinic to the claims.

16. **Risk Adjustment**

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

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

    ☐ Yes ☐ No ☒ Not applicable

    16c. **Social risk adjustment:**

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

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

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

    16d. **Risk adjustment summary:**

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

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

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

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

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

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

VALIDITY: TESTING

17. **Validity testing level: ☒ Measure score ☐ Data element ☐ Both**

18. **Method of establishing validity of the measure score:**

☒ Face validity

☒ Empirical validity testing of the measure score

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

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

Submission document: Testing attachment, section 2b2.2

Construct Validity

- The developer explored whether Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment (IET) was positively correlated with the Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence measure (FUA) and whether the two indicators within the Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment measure were positively correlated with each other.
- A Pearson correlation test was used to estimate the strength of the associations. The magnitude of correlation ranges from -1 to +1.
- For the purposes of this analysis and the intended use of this measure to evaluate the quality of care for members across health plans, correlation was considered high (strong) if the correlation coefficient is 0.75 to 1, moderate if 0.25 to 0.75, and low (weak) if 0 to 0.25.
- A p-value threshold 0.05 was used to determine the significance of the correlation coefficient.

Face Validiy

- Face validity was assessed by the standardized process of the HEDIS measure life cycle.

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

Submission document: Testing attachment, section 2b2.3

Construct Validity

Correlation between Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment and Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence in Commercial Plans – HEDIS 2018

Measure/Measure Element	**Pearson Correlation Coefficients**			
	FUA: 7 Day Indicator	**FUA: 30 Day Indicator**	**IET: Initiation Indicator**	**IET: Engagement Indicator**
IET: Initiation Indicator	0.19 P value: 0.0008	0.16 P value: 0.005	1	0.51 P value: <.0001
IET: Engagement Indicator	0.31 P value: <.0001	0.31 P value: <.0001	0.51 P value: <.0001	1

Correlations between Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment and Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence in Medicare Plans – HEDIS 2018

Measure/Measure Element	**Pearson Correlation Coefficients**			
	FUA: 7 Day Indicator	**FUA: 30 Day Indicator**	**IET: Initiation Indicator**	**IET: Engagement Indicator**
IET: Initiation Indicator	0.24 P value: .0001	0.26 P value: <.0001	1	0.59 P value: <.0001
IET: Engagement Indicator	0.39 P value: <.0001	0.41 P value: <.0001	0.59 P value: <.0001	1

Correlations between Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment and Follow-Up After Emergency Department Visit for Alcohol and other Drug Abuse or Dependence in Medicaid Plans – HEDIS 2018

Measure/Measure Element	**Pearson Correlation Coefficients**			
	FUA: 7 Day Indicator	**FUA: 30 Day Indicator**	**IET: Initiation Indicator**	**IET: Engagement Indicator**
IET: Initiation Indicator	0.13 P value: 0.10 n.s.	0.08 P value: .31 n.s.	1	0.56 P value: <.0001
IET: Engagement Indicator	0.57 P value: <.0001	0.60 P value: <.0001	0.56 P value: <.0001	1

- Most Pearson Correlation Coefficients demonstrate a significant moderate correlation.
    - One notable exception was seen in the Medicaid population; only engagement in treatment had significant positive correlation with follow-up.
- Results generally confirm the developer's hypothesis that health plans with high rates of follow-up also have high rates of initiation and engagement in treatment.

Face Validity

- Results from multiple multi-stakeholder measurement advisory panels, as well as those submitting to public comment, indicate that the measure as specified has sufficient face validity and will accurately differentiate quality across providers.
- Updates made to improve face validity (inclusion of pharmacotherapy and telehealth).

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

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

22. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

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

    ☐ **Yes**

    ☐ **No**

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

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

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

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

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

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

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

- Empirical validity testing results as well as the results from face validity testing indicate moderate validity.

ADDITIONAL RECOMMENDATIONS

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

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

2a1. Reliability – Specifications
Comments:
**Reliability looks fine and data elements are clearly defined. Could be made stronger if the measure indicated that patients had both psychosocial plus medication assisted treatment.
**No concerns
**No concerns. The measures was updated to include medications and telemedicine which is an improvement.
**The specifications are detailed enough for large scale implementation. Some changes were made to include telehealth and pharmacotherapy, as well as extending the time fram fro engagement from 30 to 34 days.
**No concerns would be reliable.

2a2. Reliability – Testing
Comments:
**None
**No concerns.
**No. Signal to noise ratios were appropriately used and interpreted.
**None. Scores were within normal range.
**No.

2b1. Validity –Testing
2b4-7. Threats to Validity
2b4. Meaningful Differences
Comments:
** none.
** No concerns.
** Validity testing seemed weak however there is a large literature showing that IET correlates with outcomes including death.
** Scores comparing IET initiation with the FUA 7 and 30 day indicator were not statistically significant for Medicaid Plans. Adequate face validity.
** No.

** None.
** Missing data is not a threat.
** No.
** The inclusion of telehealth and pharmacotherapy makes sense. However, having the numerator include all of these modalities will not allow for comparisons across modalities.
** No concerns.

2b2-3. Other Threats to Validity

2b2. Exclusions

2b3. Risk Adjustment

Comments:

** None.
**No concerns.
**No
**None.
** I am unclear as to why there is a 60 day exclusion. Would seem if a patient was in treatment within the 60 days and then has an inpatient service there is even more need to ensure follow up based on their relapse.

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 in defined fields in a combination of electronic sources.
- Data elements are coded by someone other than the individual obtaining the original information.
- In addition to the HEDIS audit, NCQA provides a system to allow "real-time" feedback from measure users.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery and claims generation?
- Are there any feasibility concerns regarding this measure?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

**This is feasible though it is possible that claims data will not accurately capture all encounters.
**This measure needs to be incentivized since population is small but performance gap is very large and effects of non-performance are significant.
**It has been collected for many years and is feasible.
**All data elements are available in electronic forms. Measure is already being used extensively by NCQA.
**No concerns.

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

- The measure is currently used for both public reporting and quality improvement and is included in the following programs: Medicaid Adult Core Set; Merit Based Incentive Payment System (MIPS) Quality Payment Program (QPP); Health Insurance Exchange Quality Rating System (QRS):Qualified Health Plan (QHP) issuers and Multi-State Plan (MSP). This measure is also reported annually in the State of Health Care Annual Report and on NCQA's website via the Health Plan Ratings/Report Cards. This measure is used in scoring for accreditation of Medicare Advantage Health Plans and used in the Quality Compass tool.

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

- Health plans that report HEDIS calculate their rates and know their performance when submitting to NCQA.
- NCQA publicly reports rates across all plans and also creates benchmarks to help plans understand how they perform relative to other plans.
- The developer reevalutes measures regularly and provides an opportunity for users to provide input.
    - Feeback has included minor clarification of specifications and general support.

Additional Feedback:

N/A

Questions for the Committee:

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

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

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

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

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

Improvement results

- Over the past three years, this measure has shown slight improvement across health plans although there is still a significant gap.
- Starting in 2018, data was stratified by diagnosis cohort (i.e., alcohol, opioid, or other drug)
    - For the initiation indicator, higher performance was seen among members with a diagnosis of opioid abuse and dependence than members with diagnoses of alcohol or other drug abuse and dependence.
    - For the engagement indicator, performance was about 10 percentage points lower for Medicare than what is observed in the Medicaid and commercial products

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

Unexpected findings (positive or negative) during implementation

The developer did not identify any unexpected findings during implementation of this measure.

Potential harms

- The developer conducts audits to verify that HEDIS specification are met and limit potential data collection and calculation method variance.

Additional Feedback:

- The following comment was submitted through NQF's QPS platform in November 2018.
    - "It's an excellent measure. However, it is limited by losing a significant amount of relevant data because it excludes multiple ASAM residential treatment levels of care. Many states and other entities would benefit greatly with more accurate data if that observation was considered by the reviewing committee to include residential levels of care in the next update of that measure.**"**

***Questions for the Committee**:*

- Do the value sets appear to be missing ASAM residential treatment levels? Any other levels of care or other factors germane to the measure missing?

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

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

4a1. Use - Accountability and Transparency
Comments:
**Face validity looks good.
**Use is still very low. This reviewer has some concern about lack of age group stratification given the scientific evidence during adolescence.
**Yes, it can be used to improve the quality of SUD treatment.
**The measure is currently being used by a number of Medicaid and Medicare programs, as well as by NCQA. One user suggested that the measure should also include residential levels of care.

**Had appropriate feedback.
4b1. Usability – Improvement
Comments:
** No unintended consequences, little harm. but unclear if the benefit is long-term.
**No concerns
**No concerns
**The benefits outway the harm for use of this measure. Improvement using the measure has already been shown between plan types and diagnostic categories.
**No issues.

Criterion 5: Related and Competing Measures

Related or competing measures
The developer did not note any related and competing measures. Staff identified the following related measures:
Related
2599: Alcohol Screening and Follow-up for People with Serious Mental Illness
3312: Continuity of Care for Medicaid Beneficiaries after Detoxification (Detox) From Alcohol and/or Drugs
2605: Follow-Up After Emergency Department Visit for Mental Illness or Alcohol and Other Drug Abuse or Dependence
2152: Preventive Care and Screening: Unhealthy Alcohol Use: Screening & Brief Counseling
Harmonization
Does the committee feel the measure application has considered other similar measures sufficiently (e.g., with the reliability tests or in the general engineering of the measure)?

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

5. Related and Competing
Comments
**Several measures exist aimed at those with SMI diagnosis. This measure is aimed at general population identified with recent substance abuse or dependence and as such makes a contribution.
** No concerns.
**No.
** While there are a number of measures that are related, they are population, plan, or service specfic measures.
** No issues - comment it is good that they included a telehealth visit as counting as a visit.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/22/2019
Public Comment
**It's an excellent measure. However, it is limited by losing a significant amount of relevant data because it excludes multiple ASAM residential treatment levels of care. Many states and other entities would benefit greatly with more accurate data if that observation was considered by the reviewing committee to include residential levels of care in the next update of that measure.
Support/Non-Support
There have been no comments or support/non-support choices as of this date.

Brief Measure Information

NQF #: 0004

Corresponding Measures:

De.2. Measure Title: Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment

Co.1.1. Measure Steward: National Committee for Quality Assurance

De.3. Brief Description of Measure: This measure assesses the degree to which the organization initiates and engages members identified with a need for alcohol and other drug (AOD) abuse and dependence services and the degree to which members initiate and continue treatment once the need has been identified. Two rates are reported:

- Initiation of AOD Treatment. The percentage of adolescent and adult members with a new episode of AOD abuse or dependence who initiate treatment through an inpatient AOD admission, outpatient visit, intensive outpatient encounter, partial hospitalization, telehealth or medication assisted treatment (MAT) within 14 days of the diagnosis.
- Engagement of AOD Treatment. The percentage of adolescent and adult members with a new episode of AOD abuse or dependence who initiated treatment and who had two or more additional AOD services or MAT within 34 days of the initiation visit.

1b.1. Developer Rationale: This measure assesses the degree to which the organization initiates and engages members identified with a need for alcohol and other drug dependence (AOD) services. By providing data on access to AOD dependence treatment across care settings, this measure provides insight on how plans and their providers may need to target education efforts and assists patient in accessing care.

S.4. Numerator Statement: Initiation of AOD Treatment:

Initiation of treatment through an inpatient AOD admission, outpatient visit, intensive outpatient encounter or partial hospitalization, telehealth or medication treatment within 14 days of the diagnosis.

Engagement of AOD Treatment:

Initiation of AOD treatment and two or more additional AOD services or medication treatment within 34 days of the initiation visit.

S.6. Denominator Statement: Patients age 13 years of age and older as of December 31 of the measurement year who were diagnosed with a new episode of alcohol or other drug dependency (AOD) during the first 10 and ½ months of the measurement year (e.g., January 1-November 15).

S.8. Denominator Exclusions: Exclude members who had a claim/encounter with a diagnosis of AOD abuse or dependence (AOD Abuse and Dependence Value Set), AOD medication treatment (AOD Medication Treatment Value Set) or an alcohol or opioid dependency treatment medication dispensing event (Medication Treatment for Alcohol Abuse or Dependence Medications List; Medication Treatment for Opioid Abuse or Dependence Medications List) during the 60 days (2 months) before the IESD.

Exclude patients who use hospice services or elect to use a hospice benefit any time during the measurement year, regardless of when the services began.

De.1. Measure Type: Process

S.17. Data Source: Claims

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Aug 10, 2009 **Most Recent Endorsement Date:** Feb 08, 2016

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A

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

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

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

0004_IET_Evidence_Form.docx

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

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

Yes

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** #0004

Measure Title: Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment

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

Date of Submission: 11/1/2018

Instructions

- *Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.*
- *Complete EITHER 1a.2, 1a.3 or 1a.4 as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Outcome: 3 Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: 5 a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured structure leads to a desired health outcome.
- Efficiency: 6 evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Outcome:

☐ Patient-reported outcome (PRO):

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

☐ Intermediate clinical outcome (*e.g., lab value*):

☒ Process: This measure assesses the degree to which the organization initiates and engages members identified with a need for alcohol and other drug (AOD) abuse and dependence services and the degree to which members initiate and continue treatment once the need has been identified.

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

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

The intended result of this process measure is to identify members with diagnosed substance abuse and dependence and assess if they initiate treatment, including Medication Assisted Treatment (MAT), within 14 days of their diagnosis and engage in ongoing care within 34 days of initiation.

The presumed pathway from process to outcomes is as follows:

1. Patient (13 years or older) diagnosed with substance abuse or dependence.
2. Patient initiates treatment through an inpatient AOD admission, outpatient visit, intensive outpatient encounter, partial hospitalization, telehealth or medication assisted treatment (MAT) within 14 days of the diagnosis
3. Patient completes two or more additional AOD treatment services or MAT within 34 days of the initiation visit.
4. Patient successfully engages in treatment (intermediate step), which supports a pathway to treatment completion and substance abuse and dependence recovery or appropriate ongoing management (desired outcome).

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

N/A

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

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

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Table 1: Clinical Practice Guideline 1

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Practice Guideline for the Treatment of Patients with Substance Use Disorders: Second Edition** • **American Psychiatric Association** • **2006** • **Work Group on Substance Use Disorders, Kleber H.D., R.D. Weiss, R.F. Anton, B.J. Rounsaville, T.P. George, E.C. Strain, S.F. Greenfield, D.M. Ziedonis, T.R. Kosten, G. Hennessy, C.P. O'Brien, H.S. Connery HS, American Psychiatric Association Steering Committee on Practice Guidelines, McIntyre J.S., S.C. Charles, D.J. Anzia, J.E. Nininger, I.A. Cook, P. Summergrad, M.T. Finnerty, S.M. Woods, B.R. Johnson, J. Yager, R. Pyles, L. Lurie, C.D. Cross, R.D. Walker, R. Peele, M.A. Barnovitz, S.H. Gray, J.P. Shemo, S. Saxena, T. Tonnu, R. Kunkle, A.B. Albert, L.J. Fochtmann, C. Hart, D. Regier. (2006). *Treatment of patients with substance use disorders, second edition.* American Psychiatric Association. Am J Psychiatry 163(8 Suppl):5-82.** • **https://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/substanceuse.pdf**

Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**2. Psychiatric management ([I]Recommended with substantial clinical confidence)** "Psychiatric management is the foundation of treatment for patients with substance use disorders [I]. Psychiatric management has the following specific objectives: motivating the patient to change, establishing and maintaining a therapeutic alliance with the patient, assessing the patient's safety and clinical status, managing the patient's intoxication and withdrawal states, developing and facilitating the patient's adherence to a treatment plan, preventing the patient's relapse, educating the patient about substance use disorders, and reducing the morbidity and sequelae of substance use disorders. Psychiatric management is generally combined with specific treatments carried out in a collaborative manner with professionals of various disciplines at a variety of sites, including community-based agencies, clinics, hospitals, detoxification programs, and residential treatment facilities. Many patients benefit from involvement in self-help group meetings, and such involvement can be encouraged as part of psychiatric management." **3. Specific treatments** "The specific pharmacological and psychosocial treatments reviewed below are generally applied in the context of programs that combine a number of different treatment modalities." **a) Pharmacological treatments ([I]Recommended with substantial clinical confidence)** "Pharmacological treatments are beneficial for selected patients with specific substance use disorders [I]. The categories of pharmacological treatments are 1) medications to treat intoxication and withdrawal states, 2) medications to decrease the reinforcing effects of abused substances, 3) agonist maintenance therapies, 4) antagonist therapies, 5) abstinence-promoting and relapse prevention therapies, and 6) medications to treat comorbid psychiatric conditions." **b) Psychosocial treatments (All [I]Recommended with substantial clinical confidence)** "Psychosocial treatments are essential components of a comprehensive treatment program [I]. Evidence-based psychosocial treatments include cognitive-behavioral therapies (CBTs, e.g., relapse prevention, social skills training), motivational enhancement therapy (MET), behavioral therapies (e.g., community reinforcement, contingency management), 12-step facilitation (TSF), psychodynamic therapy/interpersonal therapy (IPT), self-help manuals, behavioral self-control, brief interventions, case management, and group, marital, and family therapies. There is evidence to support the efficacy of integrated treatment for

	patients with a co-occurring substance use and psychiatric disorder; such treatment includes blending psychosocial therapies used to treat specific substance use disorders with psychosocial treatment approaches for other psychiatric diagnoses (e.g., CBT for depression)." **Alcohol Use Disorder** **Pharmacological Treatments (All [I]Recommended with substantial clinical confidence or II] Recommended with moderate clinical confidence):** "Specific pharmacotherapies for alcohol-dependent patients have well-established efficacy and moderate effectiveness. Naltrexone may attenuate some of the reinforcing effects of alcohol [I], although data on its long-term efficacy are limited. The use of long-acting, injectable naltrexone may promote adherence, but published research is limited and FDA approval is pending. Acamprosate, a γ-aminobutyric acid (GABA) analog that may decrease alcohol craving in abstinent individuals, may also be an effective adjunctive medication in motivated patients who are concomitantly receiving psychosocial treatment [I]. Disulfiram is an effective adjunct to a comprehensive treatment program for reliable, motivated patients whose drinking may be triggered by events that suddenly increase alcohol craving [II]." **NOTE: Please see below for APA 2017 clinical practice guideline on pharmacological treatment for alcohol use disorder.** **Psychosocial Treatments:** "Psychosocial treatments found effective for some patients with an alcohol use disorder include MET [I], CBT [I], behavioral therapies [I], TSF [I], marital and family therapies [I], group therapies [II], and psychodynamic therapy/IPT [III]. Recommending that patients participate in self-help groups, such as Alcoholics Anonymous (AA), is often helpful [I]." **Opioid Use Disorder** **Pharmacological Treatments (All [I]Recommended with substantial clinical confidence):** "Maintenance treatment with methadone or buprenorphine is appropriate for patients with a prolonged history (>1 year) of opioid dependence [I]. The goals of treatment are to achieve a stable maintenance dose of opioid agonist and facilitate engagement in a comprehensive program of rehabilitation [I]. Maintenance treatment with naltrexone is an alternative strategy [I], although the utility of this strategy is often limited by lack of patient adherence and low treatment retention." **Psychosocial Treatments:** "Psychosocial treatments are effective components of a comprehensive treatment plan for patients with an opioid use disorder [II]. Behavioral therapies (e.g., contingency management) [II], CBTs [II], psychodynamic psychotherapy [III], and group and family therapies [III] have been found to be effective for some patients with an opioid use disorder. Recommending regular participation in self-help groups may also be useful [III]."

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Authors did not specifically grade the evidence used to inform each recommendation statement. However, they provided a grading system for each individual reference cited throughout their guideline (below) based on the type of clinical study included as a supporting document. "The following coding system is used to indicate the nature of the supporting evidence in the summary recommendations and references: [A] Double-blind, randomized clinical trial. A study of an intervention in which subjects are prospectively followed over time; there are treatment and control groups; subjects are randomly assigned to the two groups; both the subjects and the investigators are blind to the assignments. [A] Randomized clinical trial. Same as above but not double-blind. [B] Clinical trial. A prospective study in which an intervention is made and the results of that intervention are tracked longitudinally; study does not meet standards for a randomized clinical trial. [C] Cohort or longitudinal study. A study in which subjects are prospectively followed over time without any specific intervention. [D] Case-control study. A study in which a group of patients is identified in the present and information about them is pursued retrospectively or backward in time. [E] Review with secondary data analysis. A structured analytic review of existing data, e.g., a meta-analysis or a decision analysis. [F] Review. A qualitative review and discussion of previously published literature without a quantitative synthesis of the data. [G] Other. Textbooks, expert opinion, case reports, and other reports not included above."
Provide all other grades and definitions from the evidence grading system	See "grade assigned to the evidence associated with the recommendation with the definition of the grade" for information about each article reviewed that met inclusion criteria for this guideline.

Grade assigned to the **recommendation** with definition of the grade	"Each recommendation is identified as meriting one of three categories of endorsement, based on the level of clinical confidence regarding the recommendation, as indicated by a bracketed Roman numeral after the statement." **Recommendation 2:** [I]Recommended with substantial clinical confidence. **Recommendation 3a (Pharmacologic Treatments):** [I]Recommended with substantial clinical confidence. **Recommendation 3b (Psychosocial Treatments):** [I]Recommended with substantial clinical confidence. Further broken down by diagnosis: **Alcohol Use Disorder: Pharmacological Treatments** (All [I]Recommended with substantial clinical confidence or II] Recommended with moderate clinical confidence) **Alcohol Use Disorder: Psychosocial Treatments:** [I]Recommended with substantial clinical confidence or [III] May be recommended on the basis of individual circumstances. **Opioid Use Disorder: Pharmacological Treatments** (All [I]Recommended with substantial clinical confidence) **Opioid Use Disorder: Psychosocial Treatments:** [I]Recommended with substantial clinical confidence), [II] Recommended with moderate clinical confidence, or [III] May be recommended on the basis of individual circumstances.
Provide all other grades and definitions from the recommendation grading system	None.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	• Authors included 1,063 studies that met inclusion criteria for this guideline after reviewing 89,231 references populated using a structured literature search in PubMed. • "[Authors completed] A comprehensive literature review to identify all relevant randomized clinical trials as well as less rigorously designed clinical trials and case series when evidence from randomized trials was unavailable." For additional details about the types of studies included as citations for this guideline, see "grade assigned to the evidence associated with the recommendation with the definition of the grade."
Estimates of benefit and consistency across studies	Across included studies, guidelines for the treatment of those with substance use disorders agree that psychosocial care, and in many cases, also pharmacological treatments, are an effective way to reduce morbidity and mortality.
What harms were identified?	N/A
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	No. The conclusions drawn from this systematic review remain relevant and current, except as superseded by more recent guidance below specific to alcohol use disorder.

Table 2: Clinical Practice Guideline 2

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **Practice Guideline for the Pharmacological Treatment of Patients with Alcohol Use Disorder** • **American Psychiatric Association** • **2018** • **Reus, V. et al. (2018). Practice Guideline for the Pharmacological Treatment of Patients with Alcohol Use Disorder. *American Journal of Psychiatry, 175(1), 86-90. doi:*10.1176/appi.ajp.2017.1750101** • **https://psychiatryonline.org/doi/pdf/10.1176/appi.books.9781615371969**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	"Statement 8. APA recommends (1C) that patients with alcohol use disorder have a documented comprehensive and person-centered treatment plan that includes evidence-based nonpharmacological and pharmacological treatments." "Statement 9. APA recommends (1B) that naltrexone or acamprosate be offered to patients with moderate to severe alcohol use disorder who • have a goal of reducing alcohol consumption or achieving abstinence, • prefer pharmacotherapy or have not responded to nonpharmacological treatments alone, and • have no contraindications to the use of these medications." "Statement 10. APA suggests (2C) that disulfiram be offered to patients with moderate to severe alcohol use disorder who • have a goal of achieving abstinence, • prefer disulfiram or are intolerant to or have not responded to naltrexone and acamprosate, • are capable of understanding the risks of alcohol consumption while taking disulfiram, and • have no contraindications to the use of this medication." "Statement 11. APA suggests (2C) that topiramate or gabapentin be offered to patients with moderate to severe alcohol use disorder who • have a goal of reducing alcohol consumption or achieving abstinence, • prefer topiramate or gabapentin or are intolerant to or have not responded to naltrexone and acamprosate, • and • have no contraindications to the use of these medications."

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Statement 8: "A" rating for evidence: High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect. Statement 9: "B" rating for evidence: Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate. Statement 10: "C" rating for evidence: Low confidence that the evidence reflects the true effect. Further research is likely to change our confidence in the estimate of effect and is likely to change the estimate. Statement 11: "C" rating for evidence: Low confidence that the evidence reflects the true effect. Further research is likely to change our confidence in the estimate of effect and is likely to change the estimate.
Provide all other grades and definitions from the evidence grading system	N/A
Grade assigned to the **recommendation** with definition of the grade	Statement 8 and Statement 9: "1" Recommendation: APA recommends with confidence that the benefits of the intervention clearly outweigh harms. Statement 10 and Statement 11: "2" Suggestion: APA suggests the that although the benefits of the statement are still viewed as outweighing the harms, the balance of benefits and harms is more difficult to judge, or either the benefits or the harms may be less clear. With a suggestion, patient values and preferences may be more variable, and this can influence the clinical decision that is ultimately made.
Provide all other grades and definitions from the recommendation grading system	N/A
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	The Agency for Healthcare Research and Quality (AHRQ) systematic review "Pharmacotherapy for Adults With Alcohol-Use Disorders in Outpatient Settings" is the source of evidence used for the development of this guideline. This systematic review included 95 randomized clinical trials, accounting for 22,803 patients. Jonas, D.E., Amick, H.R., Feltner, C., et al. (2014). Pharmacotherapy for Adults With Alcohol Use Disorders in Outpatient Settings A Systematic Review and Meta-analysis. JAMA, 311(18), 1889–1900. doi:10.1001/jama.2014.3628

Estimates of benefit and consistency across studies	The following texts are directly quoted from the APA guideline and summarize the benefits of each recommendation statement as determined by clinical evidence review: **Statement 8. Evidence-Based Treatment Planning** "Development and documentation of a comprehensive treatment plan assures that the clinician has considered the available nonpharmacological and pharmacological options for treatment and has identified those treatments that are best suited to the needs of the individual patient, with a goal of improving overall outcome. It may also assist in forming a therapeutic relationship, eliciting patient preferences, permitting education about possible treatments, setting expectations for treatment, and establishing a framework for shared decision-making. Documentation of a treatment plan promotes accurate communication among all those caring for the patient and can serve as a reminder of prior discussions about treatment." "The potential benefits of this recommendation were viewed as far outweighing the potential harms. The level of research evidence is rated as low because no information is available on the harms of such an approach. There is also minimal research on whether developing and documenting a specific treatment plan improves outcomes as compared with assessment and documentation as usual. However, the majority of studies of pharmacotherapy for AUD included nonpharmacological treatments aimed at providing supportive counseling, enhancing coping strategies, and promoting adherence. This indirect evidence supports the benefits of comprehensive treatment planning." **Statement 9. Naltrexone or Acamprosate** "Acamprosate is associated with a small benefit on the outcomes of returning to any drinking and on the number of drinking days (moderate strength of research evidence). Naltrexone is associated with a small benefit on the outcomes of returning to any drinking, returning to heavy drinking, frequency of drinking days, and frequency of heavy drinking days (moderate strength of research evidence). Evidence is limited, but the use of long-acting injectable naltrexone may have benefits for adherence as compared with oral formulations of naltrexone. In the AHRQ meta-analysis of head to- head comparisons, neither acamprosate nor naltrexone showed superiority to the other medication in terms of return to heavy drinking (moderate strength of research evidence), return to any drinking (moderate strength of research evidence), or percentage of drinking days (low strength of research evidence). However, in the U.S. COMBINE study (but not the German PREDICT study), naltrexone was associated with better outcomes than acamprosate."

	"The potential benefits of this recommendation were viewed as far outweighing the potential harms. For both acamprosate and naltrexone, the harms of treatment were considered minimal, particularly compared with the harms of continued alcohol use, as long as there was no contraindication to the use of the medication. The positive effects of acamprosate and naltrexone were small overall, and not all studies showed a statistically significant benefit from these medications. In addition, European studies showed greater benefit of acamprosate than did U.S. studies, and naltrexone exhibited greater effect than acamprosate in the COMBINE trial. Nevertheless, the potential benefit of each medication was viewed as far outweighing the harms of continued alcohol use, particularly when nonpharmacological approaches have not produced an effect or when patients prefer to use one of these medications as an initial treatment option. In addition, it was noted that even small effect sizes may be clinically meaningful because of the significant morbidity associated with AUD. Patients with mild AUD rarely participated in clinical trials of naltrexone and acamprosate pharmacotherapy. Therefore, although they might respond to these medications, patients with mild AUD are not included in this recommendation because of the limited amount of research evidence." **Statement 10. Disulfiram** "Benefits of disulfiram on alcohol-related outcomes were not reported in the AHRQ review. However, a subsequent meta-analysis (Skinner et al. 2014) that included randomized open-label studies (low strength of research evidence) showed a moderate effect of disulfiram as compared with no disulfiram as well as compared with acamprosate, naltrexone, and topiramate. In studies where medication adherence was assured through supervised administration, the effect of disulfiram was large (Skinner et al. 2014)." "The potential benefits of this statement were viewed as likely to outweigh the harms. The strength of research evidence is rated as low because there were insufficient data from double-blind randomized controlled trials (RCTs), and the bulk of the research evidence for benefits and harms was from randomized open-label studies. With carefully selected patients in clinical trials, adverse events were somewhat greater with disulfiram. However, serious adverse events were few and comparable in numbers to serious adverse events in comparison groups consistent with the long history of safe use of disulfiram in clinical practice. Consequently, the potential benefits of disulfiram were viewed as likely to outweigh the harms for most patients given the medium to large effect size for the benefit of disulfiram when open-label studies are considered and particularly compared with the harms of continued alcohol use. In addition, it was noted that

	even small effect sizes may be clinically meaningful because of the significant morbidity associated with AUD. The strength of the guideline statement (suggestion) was influenced both by the strength of research evidence and by patient preferences related to disulfiram as compared with other interventions."
What harms were identified?	The following texts are directly quoted from the APA guideline and summarize the harms of each recommendation statement as determined by clinical evidence review: **Statement 8. Evidence Based Treatment Planning** "The only identifiable harm from this recommendation relates to the time spent in discussion and documentation that may reduce the opportunity to focus on other aspects of the evaluation." **Statement 9. Naltrexone or Acamprosate** "The harms of acamprosate are small in magnitude, with slight overall increases in diarrhea and vomiting as compared with placebo (moderate strength of research evidence). The harms of naltrexone are small in magnitude, with slight overall increases in dizziness, nausea, and vomiting relative to placebo (moderate strength of research evidence). Alterations in hepatic function are also possible with naltrexone, but changes in liver chemistries were not assessed in the AHRQ review. Individuals taking naltrexone would not be able to take opioids for pain, and other treatments for acute pain would be needed. For individuals treated with long-acting injectable naltrexone, pain or induration can occur at the injection site, and access to the medication can be an issue because of geographic- or payment-related issues. With long durations of naltrexone use, individuals lose tolerance to opioids. This can result in overdose and death if large but previously tolerated opioid doses are taken after naltrexone is discontinued. For many other potential harms, including mortality, evidence was not available or was rated by the AHRQ review as insufficient. However, withdrawals from the studies due to adverse events did not differ from placebo for acamprosate (low strength of research evidence) and were only slightly greater than placebo for naltrexone although statistically significant (moderate strength of research evidence)." **Statement 10. Disulfiram** "There were insufficient data on harms of disulfiram to conduct a meta-analysis in the AHRQ report. When randomized open-label studies were included (low strength of research evidence; Skinner et al. 2014), there was a significantly greater number of adverse events with disulfiram than with control conditions. Significant harms have been reported if alcohol-containing products are ingested concomitantly with disulfiram use."

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	N/A

Table 3: Clinical Practice Guideline 3

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **The ASAM National Practice Guideline for the Use of Medications in the Treatment of Addiction Involving Opioid Use** • **American Society of Addiction Medicine (ASAM).** • **June 1, 2015** • **American Society of Addiction Medicine (ASAM). (2015). *The ASAM National Practice Guideline for the Use of Medications in the Treatment of Addiction Involving Opioid Use.* Retrieved from: https://www.asam.org/docs/default-source/practice-support/guidelines-and-consensus-docs/asam-national-practice-guideline-supplement.pdf?sfvrsn=24** • **https://www.asam.org/docs/default-source/practice-support/guidelines-and-consensus-docs/asam-national-practice-guideline-supplement.pdf?sfvrsn=24**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	*Part 7: Psychosocial Treatment in Conjunction with Medications for the Treatment of Opioid Use Disorder* (1) "Psychosocial treatment is recommended in conjunction with any pharmacological treatment of opioid use disorder. At a minimum, psychosocial treatment should include the following: psychosocial needs assessment, supportive counseling, links to existing family supports, and referrals to community services." (2) "Treatment planning should include collaboration with qualified behavioral healthcare providers to determine the optimal type and intensity of psychosocial treatment and for renegotiation of the treatment plan for circumstances in which patients do not adhere to recommended plans for, or referrals to, psychosocial treatment." (3) "Psychosocial treatment is generally recommended for patients who are receiving opioid agonist treatment (methadone or buprenorphine)." (4) "Psychosocial treatment should be offered with oral and extended-release injectable naltrexone. The efficacy of extended-release injectable naltrexone to treat opioid use disorder has not been confirmed when it has been used as pharmacotherapy without accompanying psychosocial treatment."
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	ASAM does not provide a rating for their evidence. "These guidelines were developed using the RAND/UCLA Appropriateness Method (RAM) - a process that combines scientific evidence and clinical knowledge to determine the appropriateness of a set of clinical procedures."

Provide all other grades and definitions from the evidence grading system	N/A
Grade assigned to the **recommendation** with definition of the grade	ASAM does not provide a rating for recommendation statements.
Provide all other grades and definitions from the recommendation grading system	N/A
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	"In total, 49 guidelines were identified and 34 were ultimately included in the analysis." "The majority of existing clinical guidelines are based on systematic reviews of the literature including appropriateness criteria used in the RAM. Therefore, the aim of this exercise was not to re-review all of the research literature, but to identify within the existing clinical guidelines how they addressed common questions or considerations that clinicians are likely to raise in the course of deciding whether and how to use medications as part of the treatment of individuals with opioid use disorder."
Estimates of benefit and consistency across studies	"[Across studies included in the recommendations], patients experience improved outcomes after receiving psychosocial treatment, in both individual and group formats, from a variety of approaches. Ancillary drug addiction counseling and mutual-help programs are generally considered beneficial."
What harms were identified?	"Because lack of patient understanding and adherence may adversely affect outcomes, clinicians should make every effort to promote the patient's understanding of, and adherence to, prescribed and recommended pharmacological and psychosocial treatments. Patients should be informed of the risks, benefits, and alternatives to a particular treatment, and should be an active party to shared decision-making whenever feasible."
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	n/a

Table 4: Clinical Practice Guideline 4

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• **VA/DoD Clinical Practice Guideline for the Management of Substance Use Disorders** • **Department of Veterans Affairs and Department of Defense** • **2015** • **2009** • **Department of Veteran Affairs, Department of Defense. (2015). *VA/DoD Clinical Practice Guideline for the Management of Substance Use Disorders.* Washington DC: Department of Veterans Affairs, Department of Defense.** • **Department of Veteran Affairs, Department of Defense. (2009). *VA/DoD clinical practice guideline for management of substance use disorders (SUD).* Washington (DC): Department of Veteran Affairs, Department of Defense.** • **https://www.healthquality.va.gov/guidelines/MH/sud/VADoDSUDCPGRevised22216.pdf**

Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Recommendation 3: "For patients with a diagnosis of a substance use disorder, we suggest offering referral for specialty substance use disorder care based on willingness to engage in specialty treatment" Recommendation 5: "For patients with moderate-severe alcohol use disorder, we recommend offering one of the following medications: - Acamprosate - Disulfiram - Naltrexone- oral or extended release - Topiramate" Recommendation 7: "For patients with alcohol use disorder we recommend offering one or more of the following interventions considering patient preference and provider training/competence: - Behavioral Couples Therapy for alcohol use disorder - Cognitive Behavioral Therapy for substance use disorders - Community Reinforcement Approach - Motivational Enhancement Therapy - 12-Step Facilitation" Recommendation 8: "For patients with opioid use disorder, we recommend offering one of the following medications considering patient preferences: - Buprenorphine/naloxone - Methadone in an Opioid Treatment Program" Recommendation 11: "For patients with opioid use disorder for whom opioid agonist treatment is contraindicated, unacceptable, unavailable, or discontinued and who have established abstinence for a sufficient period of time (see narrative), we recommend offering: - Extended-release injectable naltrexone" Recommendation 24: "For patients who have initiated an intensive phase of outpatient or residential treatment, we recommend offering and encouraging ongoing systematic relapse prevention efforts or recovery support individualized on the basis of treatment response." VA/DoD 2009: Offer referral to specialty SUD care for addiction treatment if the patient: • May benefit from additional evaluation or motivational interviewing regarding his/her substance use and related problems • Has tried and been unable to change substance use on his/her own or does not respond to repeated brief intervention • Has been diagnosed with substance dependence • Has previously been treated for an alcohol or other substance use disorder

Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The VA/DoD did not grade the evidence using a separate system from the overall grading of the recommendation. For the recommendation grade, see “Grade assigned to the recommendation with definition of the grade” below.
Provide all other grades and definitions from the evidence grading system	N/A

Grade assigned to the **recommendation** with definition of the grade	Grading of Recommendations Assessment, Development and Evaluation (GRADE) system to assess the quality of the evidence base and assign a grade for the strength VA/DoD Clinical Practice Guideline for the Management of Substance Use Disorders December 2015 Page 11 of 169 for each recommendation. The following grade assignments were used: • Strong For (or "We recommend offering this option …") • Weak For (or "We suggest offering this option …") • Weak Against (or "We suggest not offering this option …") • Strong Against (or "We recommend against offering this option …") The relative strength of the recommendation is based on a binary scale, "Strong" or "Weak." A strong recommendation indicates that the Work Group is highly confident that desirable outcomes outweigh undesirable outcomes. If the Work Group is less confident of the balance between desirable and undesirable outcomes, they present a weak recommendation. Similarly, a recommendation for a therapy or preventive measure indicates that the desirable consequences outweigh the undesirable consequences. A recommendation against a therapy or preventive measure indicates that the undesirable consequences outweigh the desirable consequences. Grading by Recommendation • Recommendation 3: Weak for (We suggest offering this option) • Recommendation 5: Strong for (We recommend offering this option) • Recommendation 7: Strong for (We recommend offering this option) • Recommendation 8: Strong for (We recommend offering this option) • Recommendation 11: Strong for (We recommend offering this option) • Recommendation 24: Strong for (We recommend offering this option) **A** A strong recommendation that the clinicians provide the intervention to eligible patients. Good evidence was found that the intervention improves important health outcomes and concludes that benefits substantially outweigh harm. **B** A recommendation that clinicians provide (the service) to eligible patients. At least fair evidence was found that the intervention improves health outcomes and concludes that benefits outweigh harm.

Provide all other grades and definitions from the recommendation grading system	**C** No recommendation for or against the routine provision of the intervention is made. At least fair evidence was found that the intervention can improve health outcomes, but concludes that the balance of benefits and harms is too close to justify a general recommendation. **D** Recommendation is made against routinely providing the intervention to asymptomatic patients. At least fair evidence was found that the intervention is ineffective or that harms outweigh benefits. **I** The conclusion is that the evidence is insufficient to recommend for or against routinely providing the intervention. Evidence that the intervention is effective is lacking, or poor quality, or conflicting, and the balance of benefits and harms cannot be determined.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Overall, 135 studies, the majority of which are randomized control trials or systematic reviews, were included in the systematic review used to inform this guideline.
Estimates of benefit and consistency across studies	Overall, the authors of this guideline put forth the above recommendations and specifically stated that the benefits of the recommended treatments and protocol outweigh their potential harms. Additionally, the authors discuss not only the benefit for the primary outcome of interest, engaging patients in SUD care, but the improvement in secondary outcomes, such as crime associated with substance use, social engagement and vocational productivity, transmittable diseases, and morbidity.
What harms were identified?	Overall, the authors felt that the benefit of treatment, in accordance with the recommendations put forth in the guidelines, outweighed any potential risk. For each of the pharmacotherapies discussed in the guideline, the authors explicitly urge providers to carefully consider the risks and benefits for each individual patient being treated. With regard to treatment of pregnant women, the authors included the following: "Clinicians should weigh the unknown risks of long-term harm to the fetus from limited exposure to naloxone in the combination product [buprenorphine/naloxone combination product] versus the risks of misuse or diversion posed by prescribing the mono-product to the mother during pregnancy."
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	n/a

---

1a.4 OTHER SOURCE OF EVIDENCE

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

N/A

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

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

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

1b. Performance Gap

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

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

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

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

This measure assesses the degree to which the organization initiates and engages members identified with a need for alcohol and other drug dependence (AOD) services. By providing data on access to AOD dependence treatment across care settings, this measure provides insight on how plans and their providers may need to target education efforts and assists patient in accessing care.

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

Initiation:

Medicaid

Measurement Year:	**2016;**	**2017;**	**2018***
AVE:	38.24	40.87	42.28
N:	161	184	186
Min:	13.46	14.08	11.36
Max:	54.79	65.25	64.63
SD:	6.51	7.75	7.43
P10:	30.24	31.96	33.72
P25:	34.39	35.79	38.62
P50:	38.07	40.72	42.22
P75:	42.81	45.13	46.40
P90:	46.28	50.00	50.20

*Alcohol Diagnosis Stratification:

AVE: 40.96

N: 180

Min: 15.70

Max: 62.73

SD: 6.74

P10: 33.88

P25: 36.89

P50: 40.67

P75: 44.24

P90: 48.46

*Opioid Diagnosis Stratification:

AVE: 50.25

N: 173

Min: 19.64

Max: 74.71

SD: 11.52

P10: 34.94

P25: 40.99

P50: 50.84

P75: 58.62

P90: 65.22

*Other drug Diagnosis Stratification:

AVE: 42.46

N: 183

Min: 8.33

Max: 66.65

SD: 8.96

P10: 31.32

P25: 37.98

P50: 41.95

P75: 47.40

P90: 52.70

Medicare

Measurement Year:	**2016;**	**2017;**	**2018***
AVE:	33.25	33.42	34.40
N:	397	404	408
Min:	5.26	3.93	5.98
Max:	86.10	87.88	89.35
SD:	12.21	13.23	13.11

Measurement Year:	2016;	2017;	2018*
P10:	18.29	15.59	15.21
P25:	25.44	25.26	27.01
P50:	33.33	33.63	35.06
P75:	41.03	40.92	41.71
P90:	45.95	48.78	48.30

* Alcohol Diagnosis Stratification:

AVE: 39.47

N: 370

Min: 8.82

Max: 89.82

SD: 11.33

P10: 24.52

P25: 32.86

P50: 40.40

P75: 45.52

P90: 51.47

*Opioid Diagnosis Stratification:

AVE: 31.61

N: 323

Min: 2.90

Max: 94.44

SD: 16.32

P10: 10.81

P25: 18.83

P50: 30.41

P75: 42.86

P90: 51.91

*Other Drug Diagnosis Stratification:

AVE: 32.55

N: 323

Min: 3.46

Max: 80.00

SD: 15.25

P10: 9.98

P25: 21.98

P50: 33.33

P75: 42.51

P90: 51.40

Commercial

Measurement Year:	**2016;**	**2017;**	**2018***
AVE:	33.92	33.70	36.65
N:	405	401	384
Min:	13.16	15.69	12.12
Max:	54.61	78.69	83.20
SD:	5.28	6.33	7.67
P10:	27.86	27.54	29.39
P25:	30.61	30.59	33.02
P50:	34.02	33.18	35.85
P75:	36.59	35.91	39.18
P90:	40.44	40.60	42.01

*Alcohol Diagnosis Stratification:

AVE: 37.02

N: 377

Min: 19.57

Max: 80.48

SD: 7.42

P10: 29.75

P25: 33.44

P50: 36.56

P75: 39.77

P90: 43.37

*Opioid Diagnosis Stratification:

AVE: 41.76

N: 316

Min: 3.85

Max: 95.82

SD: 11.09

P10: 27.93

P25: 35.15

P50: 41.59

P75: 48.05

P90: 53.66

*Other Drug Diagnosis Stratification:

AVE: 37.79

N: 353

Min: 7.14

Max: 84.22

SD: 8.57

P10: 29.03

P25: 32.87

P50: 37.50

P75: 41.44

P90: 46.16

Engagement:

Medicaid

Measurement Year:	**2016;**	**2017;**	**2018***
AVE:	10.31	12.66	13.55
N:	163	186	188
Min:	0.00	0.00	0.00
Max:	25.33	34.04	28.27
SD:	4.80	6.32	5.89
P10:	4.42	4.82	6.05
P25:	6.92	7.98	9.11
P50:	9.79	12.36	13.69
P75:	13.20	16.25	17.74
P90:	16.95	21.31	21.40

*Alcohol Diagnosis Stratification:

AVE: 10.71

N: 182

Min: 0.78

Max: 27.27

SD: 5.00

P10: 4.17

P25: 7.14

P50: 10.86

P75: 13.52

P90: 16.17

*Opioid Diagnosis Stratification:

AVE: 22.31

N: 175

Min: 1.46

Max: 48.87

SD: 11.63

P10: 7.34

P25: 13.10

P50: 21.23

P75: 31.48

P90: 37.48

*Other Drug Diagnosis Stratifcation:

AVE: 11.66

N: 185

Min: 0.00

Max: 28.75

SD: 5.64

P10: 4.38

P25: 8.09

P50: 11.29

P75: 15.15

P90: 18.95

Medicare

Measurement Year:	**2016;**	**2017;**	**2018***
AVE:	3.14	3.52	4.21
N:	397	404	408
Min:	0.00	0.00	0.00
Max:	17.11	15.63	16.99
SD:	2.56	2.68	2.91
P10:	0.56	0.69	0.83
P25:	1.37	1.62	2.21
P50:	2.60	3.03	3.71
P75:	4.08	4.73	5.62
P90:	6.47	7.03	8.08

*Alcohol Diagnosis Stratification:

AVE: 4.57

N: 370

Min: 0.00

Max: 18.67

SD: 2.93

P10: 1.52

P25: 2.56

P50: 4.09

P75: 5.97

P90: 8.40

*Opioid Diagnosis Stratification:

AVE: 4.61

N: 323

Min: 0.00

Max: 26.60

SD: 4.33

P10: 0.50

P25: 1.55

P50: 3.38

P75: 6.28

P90: 10.23

*Other Drug Diagnosis Stratification:

AVE: 3.66

N: 323

Min: 0.00

Max: 17.50

SD: 3.43

P10: 0.00

P25: 1.00

P50: 2.80

P75: 5.62

P90: 8.33

Commercial

MeasurementYear:	**2016;**	**2017;**	**2018***
AVE:	12.65	12.09	13.40
N:	405	402	384
Min:	0.00	0.00	0.00
Max:	30.23	29.78	25.90
SD:	4.81	4.37	4.05

MeasurementYear:	2016;	2017;	2018*
P10:	7.28	6.22	8.70
P25:	9.28	9.63	10.97
P50:	12.34	12.11	13.32
P75:	15.57	15.04	15.88
P90:	18.76	17.33	18.33

*Alcohol Diagnosis Stratification:

AVE: 12.84

N: 377

Min: 0.00

Max: 29.07

SD: 4.32

P10: 7.55

P25: 10.42

P50: 12.70

P75: 15.33

P90: 17.75

*Opioid Diagnosis Stratification:

AVE: 20.74

N: 316

Min: 1.41

Max: 44.18

SD: 8.11

P10: 10.62

P25: 14.62

P50: 20.30

P75: 26.58

P90: 30.95

*Other Drug Diagnosis Stratification:

AVE: 13.29

N: 353

Min: 0.00

Max: 25.49

SD: 4.79

P10: 7.04

P25: 10.10

P50: 13.12

P75: 16.50

P90: 19.70

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

Section 1b.2 references data from the most recent three years of measurement for this measure and includes average performance, N = number of health plans, min, max, standard deviation and percentiles (and where applicable stratification by diagnosis).

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

HEDIS data are stratified by type of insurance (e.g. commercial, Medicaid, Medicare). NCQA does not currently collect performance data stratified by race, ethnicity, or language. Escarce et al. have described in detail the difficulty of collecting valid data on race, ethnicity, and language at the health plan level (Escarce, 2011). While not specified in the measure, this measure can also be stratified by demographic variables, such as race/ethnicity or socioeconomic status, in order to assess the presence of health care disparities. NCQA's Race/Ethnicity Diversity of Membership and the Language Diversity of Membership HEDIS® measures were designed to promote standardized methods for collecting these data and follow Office of Management and Budget and Institute of Medicine guidelines for collecting and categorizing race/ethnicity and language data. In addition, NCQA's Multicultural Health Care Distinction Program outlines standards for collecting, storing, and using race/ethnicity and language data to assess health care disparities. Based on extensive work by NCQA to understand how to promote culturally and linguistically appropriate services among plans and providers, we have many examples of how health plans have used HEDIS measures to design quality improvement programs to decrease disparities in care.

Escare J.J., Carreon R., Vesolovskiy G., and Lawson E.H. 2011. Collection of Race and Ethnicity Data by Health Plans Has Grown Substantially, But Opportunities Remain to Expand Efforts. Health Affairs 20(10): 1984-1991.

The measure is not stratified to detect disparities. NCQA has participated with IOM and others in attempting to include information on disparities in measure data collection. However, at the present time, this data, at all levels (claims data, paper chart review, and electronic records), is not coded in a standard manner, and is incompletely captured. There are no consistent standards for what entity (physician, group, plan, employer) should capture and report this data. While "requiring" reporting of the data could push the field forward, it has been our position that doing so would create substantial burden with inability to use the data because of its inconsistency. At the present time, we agree with the IOM report that disparities are best considered by the use of zip code analysis which has limited applicability in most reporting situations. At the health plan level, for HEDIS health plan data collection, NCQA does have extensive data related to our use of stratification by insurance status (Medicare, Medicaid and private-commercial) and would strongly recommend this process where the data base supporting the measurement includes this information. However, we believe that the measure specifications should NOT require this since the measure is still useful where the data needed to determine disparities cannot be ascertained from the data available.

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

Although HEDIS measures are not stratified by race and ethnicity, others, including the Centers for Medicare and Medicaid Services (CMS) have explored disparities related to this quality measure. The CMS Office of Minority Health, in collaboration with the RAND Corporation, began releasing national level health care quality data for different racial/ethnic groups in 2016. The findings in the Racial and Ethnic Disparities in Health Care in Medicare Advantage report include clinical care measures and patient experience measures for Medicare beneficiaries, including the IET measure. Clinical care data are reported for Medicare Advantage beneficiaries via medical records and insurance claims for hospitalizations, medical office visits, and procedures. In 2014 the IET results indicated that Asians or Pacific Islanders and Hispanics initiated treatment within 14 days of a new episode and diagnosis of AOD abuse or dependence less frequently than Whites (CMS, 2016). Overall, 19.2 percent of Asians or Pacific Islanders; 18.2 percent of Hispanics; and 29.5 percent of Whites initiated appropriate treatment.

In 2014, Asian or Pacific Islander patients and Hispanic patients with a new episode of AOD abuse or dependence and who initiated treatment were less likely than White patients to have had two or more additional services within 30 days of the initiation visit. Overall, 1.4 percent of Asian and Pacific Islanders,1.4 percent of Hispanic and 2.6 percent of Whites had two or more additional services for their new diagnosis of AOD after initiation of treatment (CMS, 2016). Conversely, Blacks (32.5 percent) were more likely than Whites (29.5 percent) to initiate treatment within 14 days of an AOD diagnosis (CMS, 2016). However, Blacks (2.6 percent) were as likely as Whites (2.6 percent) to engage in treatment (i.e., two or more additional services with a diagnosis of AOD within 30 days of the initiation of treatment), according to 2014 findings (CMS, 2016).

Centers for Medicare and Medicaid Services Office of Minority Health. 2016. Racial and Ethnic Disparities in Health Care and Medicare Advantage. Baltimore, MD.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

Behavioral Health, Behavioral Health : Alcohol, Substance Use/Abuse

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

Care Coordination, Safety

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

Children, Elderly, Populations at Risk

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

N/A

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

This is not an eMeasure **Attachment:**

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

Attachment: Attachment 0004_IET_Value_Sets.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

Added dispensing of pharmacotherapy for treatment of alcohol and opioid abuse and dependence as appropriate initiation and engagement criteria. Medication-assisted treatment (MAT), or the use of medicine in addition to psychosocial care, is a guideline-supported treatment option for those with alcohol or opioid use disorders. Adding pharmacotherapy to the measure numerator aligns the included treatment options with current guidelines and literature.

- Added "telehealth" to the denominator and numerators. Telehealth is an evidence-supported modality for the treatment of patients with substance use disorders.
- Extended the Engagement of AOD Treatment time frame to 34 days from 30 days. The slight extension of the timeframe for engagement is to allow for all FDA-approved medication treatment (particularly the long-term injectable medications, such as naltrexone) options to be dispensed or administered, if used to satisfy the engagement criteria.

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

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

Initiation of AOD Treatment:

Initiation of treatment through an inpatient AOD admission, outpatient visit, intensive outpatient encounter or partial hospitalization, telehealth or medication treatment within 14 days of the diagnosis.

Engagement of AOD Treatment:

Initiation of AOD treatment and two or more additional AOD services or medication treatment within 34 days of the initiation visit.

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

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

Index Episode Start Date. The earliest date of service for an eligible encounter during the Intake Period with a diagnosis of AOD abuse or dependence.

- For an outpatient, intensive outpatient, partial hospitalization, observation, telehealth, detoxification or ED visit (not resulting in an inpatient stay), the IESD is the date of service.
- For an inpatient stay, the IESD is the date of discharge.
- For an ED and observation visits that results in an inpatient stay, the IESD is the date of the inpatient discharge (an AOD diagnosis is not required for the inpatient stay; use the diagnosis from the ED or observation visit to determine the diagnosis cohort).
- For direct transfers, the IESD is the discharge date from the last admission (an AOD diagnosis is not required for the transfer; use the diagnosis from the initial admission to determine the diagnosis cohort).

INITIATION OF AOD TREATMENT

Initiation of AOD treatment within 14 days of the IESD.

If the Index Episode was an inpatient discharge (or an ED visit that resulted in an inpatient stay), the inpatient stay is considered initiation of treatment and the member is compliant.

If the Index Episode was not an inpatient discharge, the member must initiate treatment on the IESD or in the 13 days after the IESD (14 total days). Any of the following code combinations meet criteria for initiation:

- An acute or nonacute inpatient admission with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set. To identify acute and nonacute inpatient admissions:
- Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set).
- Identify the admission date for the stay.
- IET Stand Alone Visits Value Set with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set, with or without a telehealth modifier (Telehealth Modifier Value Set).
- Observation Value Set with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set.
- IET Visits Group 1 Value Set with IET POS Group 1 Value Set and a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set with or without a telehealth modifier (Telehealth Modifier Value Set).
- IET Visits Group 2 Value Set with IET POS Group 2 Value Set and a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set with or without a telehealth modifier (Telehealth Modifier Value Set).
- A telephone visit (Telephone Visit Value Set) with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set.
- An online assessment (Online Assessment Value) set with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set.

- If the Index Episode was for a diagnosis of alcohol abuse or dependence (Alcohol Abuse and Dependence Value Set) a medication treatment dispensing event (Medication Treatment for Alcohol Abuse or Dependence Medications List) or medication treatment during a visit (AOD Medication Treatment Value Set).
- If the Index Episode was for a diagnosis of opioid abuse or dependence (Opioid Abuse and Dependence Value Set) a medication treatment dispensing event (Medication Treatment for Opioid Abuse or Dependence Medications List) or medication treatment during a visit (AOD Medication Treatment Value Set).

For all initiation events except medication treatment (AOD Medication Treatment Value Set; Medication Treatment for Alcohol Abuse or Dependence Medications List; Medication Treatment for Opioid Abuse or Dependence Medications List), initiation on the same day as the IESD must be with different providers in order to count.

- If a member is compliant for the Initiation numerator for any diagnosis cohort (i.e., alcohol, opioid, other drug) or for multiple cohorts, count the member only once in the Total Initiation numerator. The "Total" column is not the sum of the diagnosis columns.
- Exclude the member from the denominator for both indicators (Initiation of AOD Treatment and Engagement of AOD Treatment) if the initiation of treatment event is an inpatient stay with a discharge date after November 27 of the measurement year.

ENGAGEMENT OF AOD TREATMENT

1) Numerator compliant for the Initiation of AOD Treatment numerator and
2) Members whose initiation of AOD treatment was a medication treatment event (Medication Treatment for Alcohol Abuse or Dependence Medications List; Medication Treatment for Opioid Abuse or Dependence Medications List; AOD Medication Treatment Value Set).
3) These members are numerator compliant if they have two or more engagement events where only one can be an engagement medication treatment event.
4) Remaining members whose initiation of AOD treatment was not a medication treatment event (members not identified in step 2).

These members are numerator compliant if they meet either of the following:

- At least one engagement medication treatment event.
- At least two engagement visits

Two engagement visits can be on the same date of service, but they must be with different providers in order to count as two events. An engagement visit on the same date of service as an engagement medication treatment event meets criteria (there is no requirement that they be with different providers).

Engagement visits:

Any of the following meet criteria for an engagement visit:

- An acute or nonacute inpatient admission with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set. To identify acute or nonacute inpatient admissions:
  - Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set).
  - Identify the admission date for the stay.
- IET Stand Alone Visits Value Set with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other

Drug Abuse and Dependence Value Set, with or without a telehealth modifier (Telehealth Modifier Value Set).

- Observation Value Set with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set.
- IET Visits Group 1 Value Set with IET POS Group 1 Value Set with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set, with or without a telehealth modifier (Telehealth Modifier Value Set).
- IET Visits Group 2 Value Set with IET POS Group 2 Value Set with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set, with or without a telehealth modifier (Telehealth Modifier Value Set).
- A telephone visit (Telephone Visits Value Set) with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set.
- An online assessment (Online Assessments Value Set) with a diagnosis matching the IESD diagnosis cohort using one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set.
- Engagement Medication Treatment Events:
- Either of the following meets criteria for an engagement medication treatment event:
- If the IESD diagnosis was a diagnosis of alcohol abuse or dependence (Alcohol Abuse and Dependence Value Set), one or more medication treatment dispensing events (Medication Treatment for Alcohol Abuse or Dependence Medications List) or medication treatment during a visit (AOD Medication Treatment Value Set), beginning on the day after the initiation encounter through 34 days after the initiation event (total of 34 days), meets criteria for Alcohol Abuse and Dependence Treatment.
- If the IESD diagnosis was a diagnosis of opioid abuse or dependence (Opioid Abuse and Dependence Value Set), one or more medication dispensing events (Medication Treatment for Opioid Abuse or Dependence Medications List) or medication treatment during a visit (AOD Medication Treatment Value Set), beginning on the day after the initiation encounter through 34 days after the initiation event (total of 34 days), meets criteria for Opioid Abuse and Dependence Treatment.

If the member is compliant for multiple cohorts, only count the member once for the Total Engagement numerator. The Total Column is not the sum of the diagnosis columns.

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

Patients age 13 years of age and older as of December 31 of the measurement year who were diagnosed with a new episode of alcohol or other drug dependency (AOD) during the first 10 and ½ months of the measurement year (e.g., January 1-November 15).

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

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

Identify the Index Episode. Identify all members 13 years and older as of December 31 of the measurement year who during the Intake Period had one of the following:

- An outpatient visit, telehealth, intensive outpatient visit or partial hospitalization with a diagnosis of AOD abuse or dependence. Any of the following code combinations meet criteria:
  - IET Stand Alone Visits Value Set with one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set, with or without a telehealth modifier (Telehealth Modifier Value Set).
  - IET Visits Group 1 Value Set with IET POS Group 1 Value Set and with one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set, with or without a telehealth modifier (Telehealth Modifier Value Set).
  - IET Visits Group 2 Value Set with IET POS Group 2 Value Set and with one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set, with or without a telehealth modifier (Telehealth Modifier Value Set).
- A detoxification visit (Detoxification Value Set) with one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set.
- An ED visit (ED Value Set) with one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set.
- An observation visit (Observation Value Set) with one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set.
- An acute or nonacute inpatient discharge with one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set. To identify acute and nonacute inpatient discharges:
  - Identify all acute and nonacute inpatient stays (Inpatient Stay Value Set).
  - Identify the discharge date for the stay.
- A telephone visit (Telephone Visits Value Set) with one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set.
- An online assessment (Online Assessments Value Set) with one of the following: Alcohol Abuse and Dependence Value Set, Opioid Abuse and Dependence Value Set, Other Drug Abuse and Dependence Value Set.

For members with more than one episode of AOD abuse or dependence, use the first episode.

For members whose first episode was an ED or observation visit that resulted in an inpatient stay, use the diagnosis from the ED or observation visit to determine the diagnosis cohort and use the inpatient discharge date as the IESD.

Select the Index Episode Start Date.

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

Exclude members who had a claim/encounter with a diagnosis of AOD abuse or dependence (AOD Abuse and Dependence Value Set), AOD medication treatment (AOD Medication Treatment Value Set) or an alcohol or opioid dependency treatment medication dispensing event (Medication Treatment for Alcohol Abuse or Dependence Medications List; Medication Treatment for Opioid Abuse or Dependence Medications List) during the 60 days (2 months) before the IESD.

Exclude patients who use hospice services or elect to use a hospice benefit any time during the measurement year, regardless of when the services began.

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

Exclude patients who had a claim/encounter with a diagnosis of AOD during the 60 days (2 months) before the Index Episode Start Date. (See corresponding Excel document for the AOD Dependence Value Set)

- For an inpatient Index Episode Start Date, use the admission date to determine if the patient had a period of 60 days prior to the Index Episode Start Date with no claims with a diagnosis of AOD dependence.
- For an ED visit that results in an inpatient event, use the ED date of service to determine if the patient had a period of 60 days prior to the Index Episode Start Date with no claims with a diagnosis of AOD dependence.
- For direct transfers, use the first admission to determine if the patient had a period of 60 days prior to the Index Episode Start Date with no claims with a diagnosis of AOD dependence.

Exclude from the denominator for both indicators (Initiation of AOD Treatment and Engagement of AOD Treatment) patients whose initiation of treatment event is an inpatient stay with a discharge date after December 1 of the measurement year.

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

The total population is stratified by age: 13-17 and 18+ years of age.

- Report two age stratifications and a total rate.
- The total is the sum of the age stratifications.

Report the following diagnosis cohorts for each age stratification and the total rate:

- Alcohol abuse or dependence.
- Opioid abuse or dependence.
- Other drug abuse or dependence.
- Total.

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Higher score

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

Step 1. Determine the eligible population. The eligible population is all patients who satisfy all specified denominator criteria (S7-S9).

Step 2. Search administrative systems to identify numerator events for all patients in the eligible population (S6).

Step 3. Calculate the rate of numerator events in the eligible population.

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

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

N/A

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

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

N/A

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

If other, please describe in S.18.

Claims

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

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

NCQA collects HEDIS data directly from Health Management Organizations and Preferred Provider Organizations via a data submission portal - the Interactive Data Submission System (IDSS).

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

No data collection instrument provided

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

Health Plan

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

Emergency Department and Services, Inpatient/Hospital, Outpatient Services

If other:

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

2. Validity – See attached Measure Testing Submission Form

NQF_MTF_IET.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)
Measure Number (*if previously endorsed*)**:** 0004
Measure Title: Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment
Date of Submission: 8/15/2018

Type of Measure:

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

Instructions

- 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.
- For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.
- For outcome and resource use measures, section 2b3 also must be completed.
- If specified for multiple data sources/sets of specificaitons (e.g., claims and EHRs), section 2b5 also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). *Contact NQF staff if more pages are needed.*
- Contact NQF staff regarding questions. Check for resources at 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 version 7.1 of the Measure Testing Attachment.

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.

2a2. Reliability testing [10] 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 [11] 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; [12]
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). [13]
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.
Notes
10. 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).
11. 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.
12. 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.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions.

15. 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.

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

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

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

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

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

2018 submission

N/A

2012 Submission

This measure is based on administrative claims collected in the course of providing care to health plan members. NCQA collects the Healthcare Effectiveness Data and Information Set (HEDIS) data for this measure directly from health plans via the Interactive Data Submission System (IDSS) portal.

The URL is: http://www.ncqa.org/tabid/370/default.aspx

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

2018 Submission

Testing of measure score reliability and construct validity was performed using data from 2017.

2012 submission: 2010 data

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

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

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

2018 Submission

Data for measure score reliability testing: The measure score reliability was calculated from data in calendar year 2017. The number of health plans in the sample for the Initiation indicator included 408 Medicare health plans, 186 Medicaid health plans, and 384 commercial health plans. The number of health plans in the sample for the Engagement indicator included 408 Medicare health plans, 188 Medicaid health plans, and 384 commercial health plans. The sample data included all Medicare, Medicaid and commercial health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size.

Systematic evaluation of face validity: NCQA's Committee on Performance Measurement (CPM) oversees the evolution of the measurement set and includes representation by purchasers, consumers, health plans, health care providers, and policy makers. This panel is made up of 21 members. The CPM is organized and managed by NCQA and reports to the NCQA Board of Directors and is responsible for advising NCQA staff on the development and maintenance of performance measures. CPM members reflect the diversity of constituencies that performance measurement serves; some bring other perspectives and additional expertise in quality management and the science of measurement.

Data for Construct Validity Testing: Construct validity was calculated from HEDIS 2018 data, which represents calendar year 2017. The number of health plans in the sample for the Initiation indicator included 408 Medicare health plans, 186 Medicaid health plans, and 384 commercial health plans. The number of health plans in the sample for the Engagement indicator included 108 Medicare health plans, 188 Medicaid health plans, and 384 commercial health plans. The sample data included all Medicare, Medicaid and commercial health plans submitting data to NCQA for HEDIS. The plans were geographically diverse and varied in size.

2012 Submission

HEDIS Health Plan performance data for the 2010

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

2018 Submission

Patient population for measure score reliability testing: The most recent available data indicates that for 2016, HEDIS data covered 111.5 million commercial health plan members, 53.4 million Medicaid members and 19 million Medicare beneficiaries. Data is summarized at the health plan level and stratified by product line (i.e. commercial, Medicare, Medicaid). Below is a description of the population measured for 2017. It includes

number of health plans included in the analysis and the median eligible population for the measure across health plans.

Product Line	Measure Indicator	Number of Plans	Median number of eligible patients for this measure per plan
Commercial	Initiation	384	1,617
	Engagement	384	1,617
Medicare	Initiation	408	1,328
	Engagement	408	1,328
Medicaid	Initiation	186	3,967
	Engagement	188	3,963

Patient population for Construct Validity Testing: The most recent available data indicates that for 2016, HEDIS data covered 111.5 million commercial health plan members, 53.4 million Medicaid members and 19 million Medicare beneficiaries. Data is summarized at the health plan level. Data are stratified by product line (i.e. commercial, Medicare, Medicaid). Below is a description of the measured entities that include HEDIS data collection and the median eligible population for the measure across health plans for 2017.

Product Line	Measure Indicator	Number of Plans	Median number of eligible patients for this measure per plan
Commercial	Initiation	384	1,617
	Engagement	384	1,617
Medicare	Initiation	408	1,328
	Engagement	408	1,328
Medicaid	Initiation	186	3,967
	Engagement	188	3,963

2012 Submission

HEDIS Health Plan performance data for the 2010

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

2018 Submission

Reliability of the measure score was tested using a beta-binomial calculation. This analysis included the entire HEDIS data sample (described above).

Validity was demonstrated through a systematic assessment of face validity and construct validity. Per NQF instructions, we have described the composition of the technical expert panel which assessed face validity of the measure. Construct validity was demonstrated through a correlation analysis.

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

2018 Submission

We did not analyze social risk factors. Measure performance was assessed by Medicaid, commercial and Medicare plan types, which serves as a proxy for income.

2a2. RELIABILITY TESTING

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

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

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

2018 Submission

Reliability Testing of Performance Measure Score: We utilized the Beta-binomial model (Adams 2009) to assess how well one can confidently distinguish the performance of one accountable 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. The Beta-binomial model is an appropriate model when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS measures. Reliability scores range from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (i.e., noise), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities).

Adams, J.L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

2012 Submission

In order to assess measure precision in the context of the observed variability across accountable entities, we utilized the reliability estimate proposed by Adams (2009) in work produced for the National Committee for Quality Assurance (NCQA).

The following is quoted from the tutorial which focused on provider-level assessment: "Reliability is a key metric of the suitability of a measure for [provider] profiling because it describes how well one can confidently distinguish the performance of one physician from another. Conceptually, it is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. There are three main drivers of reliability: sample size, differences between physicians, and measurement error. At the physician level, sample size can be increased by increasing the number of patients in the physician's data as well as increasing the number of measures per patient." This approach is also relevant to health plans and other accountable entities.

The beta-binomial approach accounts for the non-normal distribution of performance within and across accountable entities. Reliability scores vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise or the individual accountable entity variance), whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (across accountable entities). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. Adams' approach uses a Beta-binomial model to estimate reliability; this model provides a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® measures.

Adams, J. L. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California: RAND Corporation. TR-653-NCQA, 2009

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

2018 Submission

Beta-binomial reliability

Measure Indicator Rate	Beta Binomial Reliability		
	Commercial Product	Medicare Product	Medicaid Product
Initiation	0.97	0.99	0.99
Engagement	0.94	0.96	0.99

2012 Submission

Initiation of AOD Treatment

Commercial

Total: 0.962184

13 – 17 Years: 0.697888

18 Years and Older: 0.961216

Medicaid

Total: 0.9836665

13 – 17 Years: 0.930377

18 Years and Older: 0.983049

Medicare

Total: 0.9732890

18 Years and Older: 0.9732890

Engagement of AOD Treatment.

Total: 0.967456

13 – 17 Years: 0.788911

18 Years and Older: 0.965894

Medicaid

Total: 0.992259

13 – 17 Years: 0.961001

18 Years and Older: 0.99193

Medicare

Total: 0.872810

18 Years and Older: 0.872810

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

2018 Submission

Interpretation of measure score reliability testing:

Reliability scores can vary from 0.0 to 1.0. A score of zero implies that all variation is attributed to measurement error (noise) whereas a reliability of 1.0 implies that all variation is caused by a real difference in performance (signal). Generally, a minimum reliability score of 0.7 is used to indicate sufficient signal strength to discriminate performance between accountable entities. The testing suggests the two indicators within this measure have good reliability between 0.7 and 1.0.

_____

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

2018 submission:

Method of testing construct validity: We tested for construct validity by exploring whether *Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment* (IET) was correlated with the *Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence* measure (FUA). We also examined whether the two indicators within the Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment measure were correlated with each other. We hypothesized that organizations that perform well on the FUA measure should also perform well on the IET measure given that they are similar concepts. We also hypothesized that health plans perform well on one of the two indicators in the Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment should perform well on the other indicator because they are similar constructs.

To test these correlations, we used a Pearson correlation test. This test estimates the strength of the linear association between two continuous variables; the magnitude of correlation ranges from -1 to +1. A value of 1 indicates a perfect linear dependence in which increasing values on one variable is associated with increasing values of the second variable. A value of 0 indicates no linear association. A value of -1 indicates a perfect linear relationship in which increasing values of the first variable is associated with decreasing values of the second variable. For the purposes of this analysis and the intended use of this measure to evaluate the quality of care for members across health plans, correlation was considered high (strong) if the correlation coefficient is 0.75 to 1, moderate if 0.25 to 0.75, and low (weak) if 0 to 0.25.

The significance of a 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. We used a threshold of 0.05, as p-values less than this threshold imply it is unlikely that a non-zero coefficient was observed due to chance alone.

For this measure, we specifically hypothesized:

1. Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence (both 7 Day and 30 day follow-up indicators) will be positively correlated with Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment (both initiation and engagement indicators) (i.e. plans that have high rates of follow-up will have high rates of inhiation and engagement in treatment)
2. The Initiation and Engagement of Alcohol and Other Drug Abuse and Dependence Treatment Initiation Rate will be positively correlated with the Engagement Rate (i.e. plans that have high rates of initiation of treatment will have high rates of engagement in treatment).

Method of Assessing Face Validity: NCQA has identified and refined measure management into a standardized process called the HEDIS measure life cycle.

STEP 1: NCQA staff identifies areas of interest or gaps in care. Clinical expert panels (MAPs – whose members are authorities on clinical priorities for measurement) participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness, and feasibility. This information is gathered into a work-up format. Refer to What Makes a Measure "Desirable"? The work-up is vetted by NCQA's Measurement Advisory Panels (MAPs), the Technical Measurement Advisory Panel (TMAP) and the Committee on Performance Measurement (CPM) as well as other panels as necessary.

STEP 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase. Development includes the following tasks: (1) Prepare a detailed conceptual and operational work-up that includes a testing proposal and (2) Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures. The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

STEP 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to NCQA and the CPM about new measures or about changes to existing measures. NCQA MAPs and the technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM and NCQA's Board of Directors will be included in the next HEDIS year and reported as first-year measures.

STEP 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's State of Health Care Quality, Quality Compass or in accreditation scoring. The first-year distinction guarantees that a measure can be effectively collected, reported, and audited before it is used for public accountability or accreditation. This is not testing – the measure was already tested as part of its development – rather, it ensures that there are no unforeseen problems when the measure is implemented in the real world. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues. After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

STEP 5: Public reporting is based on the first-year measure evaluation results. If the measure is approved, it will be publicly reported and may be used for scoring in accreditation.

STEP 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reviewed for reevaluation at least every three years. NCQA staff continually monitors the performance of publicly reported measures. Statistical analysis, audit result review, and user comments through NCQA's Policy Clarification Support portal contribute to measure refinement during re-evaluation, information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, NCQA prioritizes measures for re-evaluation and selected measures are researched for changes in clinical guidelines or in the health care delivery systems, and the results from previous years are analyzed. Measure work-ups are updated with new information gathered from the literature review, and the appropriate MAPs review the work-ups and the previous year's data. If necessary, the measure specification may be updated or the measure may be recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the new year's HEDIS specification manual.

2012 Submission

NCQA identified and refined measure management into a standardized process called the HEDIS measure life cycle.

*Step 1: Topic selection is the process of identifying measures that meet criteria consistent with the overall model for performance measurement. There is a huge universe of potential performance measures for future versions of HEDIS. The first step is identifying measures that meet formal criteria for further development.

NCQA staff identifies areas of interest or gaps in care. Clinical expert panels (MAPs—whose members are authorities on clinical priorities for measurement) participate in this process. Once topics are identified, a literature review is conducted to find supporting documentation on their importance, scientific soundness and feasibility. This information is gathered into a work-up format. Refer to What Makes a Measure "Desirable"? The work-up is vetted by NCQA's MAPs, the TAG, the HEDIS Policy Panel and various other panels.

*Step 2: Development ensures that measures are fully defined and tested before the organization collects them. MAPs participate in this process by helping identify the best measures for assessing health care performance in clinical areas identified in the topic selection phase.

Development includes the following tasks.

1. Ensure funding throughout measure testing
2. Prepare a detailed conceptual and operational work-up that includes a testing proposal
3. Collaborate with health plans to conduct field-tests that assess the feasibility and validity of potential measures

The CPM uses testing results and proposed final specifications to determine if the measure will move forward to Public Comment.

*Step 3: Public Comment is a 30-day period of review that allows interested parties to offer feedback to the CPM about new measures or about changes to existing measures.

NCQA MAPs and technical panels consider all comments and advise NCQA staff on appropriate recommendations brought to the CPM. The CPM reviews all comments before making a final decision about Public Comment measures. New measures and changes to existing measures approved by the CPM will be included in the next HEDIS year and reported as first-year measures.

*Step 4: First-year data collection requires organizations to collect, be audited on and report these measures, but results are not publicly reported in the first year and are not included in NCQA's Quality Compass? or in accreditation scoring.

The first-year distinction guarantees that a measure can be efficiently collected, reported and audited before it is used for public accountability or accreditation. This is not testing—the measure was already tested as part of its development—rather, it ensures that there are no unforeseen problems when the measure is implemented in the real world. NCQA's experience is that the first year of large-scale data collection often reveals unanticipated issues.

After collection, reporting and auditing on a one-year introductory basis, NCQA conducts a detailed evaluation of first-year data. The CPM uses evaluation results to decide whether the measure should become publicly reportable or whether it needs further modifications.

*Step 5: Public reporting is based on the first-year measure evaluation results. If the measure is approved, it will be reported in Quality Compass and may be used for scoring in accreditation.

Step 6: Evaluation is the ongoing review of a measure's performance and recommendations for its modification or retirement. Every measure is reevaluated at least every three years. NCQA staff continually monitors the performance of publicly reported measures. Statistical analysis, audit result review and user comments contribute to measure evaluation. Information derived from analyzing the performance of existing measures is used to improve development of the next generation of measures.

Each year, a third of the measurement set is researched for changes in clinical guidelines or health care delivery systems, and the results from previous years are analyzed. Measure work-ups are updated with new information gathered from the literature review, and the appropriate MAPs review the work-ups and the previous year's data. If necessary, the measure specification may be updated or the measure may be

recommended for retirement. The CPM reviews recommendations from the evaluation process and approves or rejects the recommendation. If approved, the change is included in the next year's HEDIS Volume 2.

What makes a measure "Desirable"?

Whether considering the value of a new measure or the continuing worth of an existing one, we must define what makes a measure useful. HEDIS measures encourage improvement. The defining question for all performance measurement—"Where can measurement make a difference?"—can be answered only after considering many factors. NCQA has established three areas of desirable characteristics for HEDIS measures, discussed below.

1. Relevance: Measures should address features that apply to purchasers or consumers, or which will stimulate internal efforts toward quality improvement. More specifically, relevance includes the following attributes.

Meaningful: What is the significance of the measure to the different groups concerned with health care? Is the measure easily interpreted? Are the results meaningful to target audiences?

Measures should be meaningful to at least one HEDIS audience (e.g., individual consumers, purchasers or health care systems). Decision makers should be able to understand a measure's clinical and economic significance.

Important to health: What is the prevalence and overall impact of the condition in the U.S. population? What significant health care aspects will the measure address?

We should consider the type of measure (e.g., outcome or process), the prevalence of medical condition addressed by the measure and the seriousness of affected health outcomes.

Financially important: What financial implications result from actions evaluated by the measure? Does the measure relate to activities with high financial impact?

Measures should relate to activities that have high financial impact.

Cost effective: What is the cost benefit of implementing the change in the health care system? Does the measure encourage the use of cost-effective activities or discourage the use of activities that have low cost-effectiveness? Measures should encourage the use of cost-effective activities or discourage the use of activities that have low cost-effectiveness.

Strategically important: What are the policy implications? Does the measure encourage activities that use resources efficiently? Measures should encourage activities that use resources most efficiently to maximize member health.

Controllable: What impact can the organization have on the condition or disease? What impact can the organization have on the measure? Health care systems should be able to improve their performance. For outcome measures, at least one process should be controlled and have an important effect on outcome. For process measures, there should be a strong link between the process and desired outcome.

Variation across systems: Will there be variation across systems? There should be the potential for wide variation across systems.

Potential for improvement: Will organizations be able to improve performance? There should be substantial room for performance improvement.

2. Scientific soundness: Perhaps in no other industry is scientific soundness as important as in health care. Scientific soundness must be a core value of our health care system—a system that has extended and improved the lives of countless individuals.

Clinical evidence: Is there strong evidence to support the measure? Are there published guidelines for the condition? Do the guidelines discuss aspects of the measure? Does evidence document a link between clinical processes and outcomes addressed by the measure? There should be evidence documenting a link between clinical processes and outcomes.

Reproducible: Are results consistent? Measures should produce the same results when repeated in the same population and setting.

Valid: Does the measure make sense? Measures should make sense logically and clinically, and should correlate well with other measures of the same aspects of care.

Accurate: How well does the measure evaluate what is happening? Measures should precisely evaluate what is actually happening.

Risk adjustment: Is it appropriate to stratify the measure by age or another variable? Measure variables should not differ appreciably beyond the health care system's control, or variables should be known and measurable. Risk stratification or a validated model for calculating an adjusted result can be used for measures with confounding variables.

Comparability of data sources: How do different systems affect accuracy, reproducibility and validity? Accuracy, reproducibility and validity should not be affected if different systems use different data sources for a measure.

3. Feasibility:

The goal is not only to include feasible measures, but also to catalyze a process whereby relevant measures can be made feasible.

Precise specifications: Are there clear specifications for data sources and methods for data collection and reporting? Measures should have clear specifications for data sources and methods for data collection and reporting.

Reasonable cost: Does the measure impose a burden on health care systems? Measures should not impose an inappropriate burden on health care systems.

Confidentiality: Does data collection meet accepted standards of member confidentiality?

Data collection should not violate accepted standards of member confidentiality. Logistical feasibility

Are the required data available?

Auditability: Is the measure susceptible to exploitation or "gaming" that would be undetectable in an audit? Measures should not be susceptible to manipulation that would be undetectable in an audit.

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

2018 Submission

Results of construct validity testing:

The results in Table 1a describe the correlations observed for Commercial plans. The results indicate that the Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment *Initiation indicator* and the Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence 7 and 30 Day Indicators had significant weak positive correlations (0.19 and 0.16, respectively). The Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment *Engagement indicator* and the Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence 7 and 30 Day Indicators had significant moderate positive correlations (0.31 and 0.31, respectively).

The results in Table 1b describe the correlations observed for Medicare plans. The results indicate that the Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment Initiation indicator and the Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence 7 and 30 Day Indicators had significant weak positive correlations (0.24 and 0.26, respectively). The Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment Engagement indicator and the Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence 7 and 30 Day Indicators had significant moderate positive correlations (0.39 and 0.41, respectively).

The results in Table 1c describe the correlations observed for Medicaid plans. The results indicate that the Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment Initiation indicator and

the Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence 7 and 30 Day Indicators had insignificant correlations (0.13 and 0.08, respectively). The Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment Engagement indicator and the Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence 7 and 30 Day Indicators had significant moderate positive correlations (0.57 and 0.60, respectively).

The results in Tables 1a, 1b, and 1c also indicate that the Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment indicators were significantly ($p<.05$) correlated with each other in the direction that was hypothesized (positively). The level of correlations among these indicators is moderate (0.51- 0.59) across the various product lines (Medicare, Medicaid, commercial).

Table 1a. Correlation between Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment and Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence in Commercial Plans – HEDIS 2018

Measure/Measure Element	Pearson Correlation Coefficients			
	FUA: 7 Day Indicator	FUA: 30 Day Indicator	IET: Initiation Indicator	IET: Engagement Indicator
IET: Initiation Indicator	0.19 P value: 0.0008	0.16 P value: 0.005	1	0.51 P value: <.0001
IET: Engagement Indicator	0.31 P value: <.0001	0.31 P value: <.0001	0.51 P value: <.0001	1

IET: Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment

FUA: Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence

Table 1b. Correlations between Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment and Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence in Medicare Plans – HEDIS 2018

Measure/Measure Element	Pearson Correlation Coefficients			
	FUA: 7 Day Indicator	FUA: 30 Day Indicator	IET: Initiation Indicator	IET: Engagement Indicator
IET: Initiation Indicator	0.24 P value: .0001	0.26 P value: <.0001	1	0.59 P value: <.0001
IET: Engagement Indicator	0.39 P value: <.0001	0.41 P value: <.0001	0.59 P value: <.0001	1

IET: Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment

FUA: Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence

Table 1c. Correlations between Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment and Follow-Up After Emergency Department Visit for Alcohol and other Drug Abuse or Dependence in Medicaid Plans – HEDIS 2018

Measure/Measure Element	Pearson Correlation Coefficients			
	FUA: 7 Day Indicator	FUA: 30 Day Indicator	IET: Initiation Indicator	IET: Engagement Indicator
IET: Initiation Indicator	0.13 P value: 0.10	0.08 P value: .31	1	0.56 P value: <.0001
IET: Engagement Indicator	0.57 P value: <.0001	0.60 P value: <.0001	0.56 P value: <.0001	1

IET: Initiation and Engagement of Alcohol and Other Drug Abuse or Dependence Treatment

FUA: Follow-Up After Emergency Department Visit for Alcohol and Other Drug Abuse or Dependence

Results of face validity assessment:

Since the last endorsement of this measure, small updates were made to bring the measure into alignment with the most recent clinical practice guidelines and to improve the face validity of the measure. These updates include the inclusion of pharmacotherapy for the treatment of opioid and alcohol abuse and dependence and the inclusion of telehealth as an appropriate way to deliver treatment for those with substance abuse and dependence. Results from multiple multi-stakeholder measurement advisory panels, as well as those submitting to public comment, indicate that the measure as specified will accurately differentiate quality across providers and has sufficient face validity.

2012 Submission

Step 1: The Initiation and Engagement of Alcohol and Other Drug Dependence measure was developed to address a gap in care concerning follow-up care for people with alcohol or other drug dependence. NCQA's Performance Measurement Department, the Behavioral Health MAP and The Washington Circle worked together to assess the most appropriate tools for monitoring follow-up for AOD.

Step 2: The measure was written, field-tested, and presented to the CPM in 2004. The CPM recommended to send the measure to public comment with a vote of 14 in favor and none opposed.

Step 3: The measure was released for Public Comment in spring 2004. We received and responded to comments on this measure. The CPM recommended moving this measure to first year data collection with a vote of 14 in favor and none opposed.

Step 4: The Initiation and Engagement of Alcohol and Other Drug Dependence measure was introduced in HEDIS 2005. Organizations reported the measures in the first year and the results were analyzed for public reporting in the following year. The CPM recommended moving this measure public reporting with a vote of 16 in favor and none opposed.

Step 5: The Initiation and Engagement of Alcohol and Other Drug Dependence measure was reevaluated in 2011/2012.

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

2018 Submission

Results of face validity assessment:

Results from multiple multi-stakeholder measurement advisory panels, as well as those submitting to public comment, indicate that the measure as specified will accurately differentiate quality across providers and has sufficient face validity.

Interpretation of construct validity testing: The results confirmed the hypothesis that health plans with high rates of follow-up also have high rates of initiation and engagement in treatment (exception seen in the Medicaid population; only engagement in treatment had significant positive correlation with follow-up). The results also confirmed the hypothesis that the Initiation and Engagement measure indicators are correlated with each other, suggesting they represent the same underlying quality construct of substance abuse and dependence care. These results indicate that the Initiation and Engagement measure is a valid measure of a plan's quality of managing substance abuse or dependence treatment.

_____

2b2. EXCLUSIONS ANALYSIS

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

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

Testing was not performed for exclusions.

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

Testing was not performed for exclusions.

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Testing was not performed for exclusions.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

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

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

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

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

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

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

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

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

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

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

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

If stratified, skip to 2b3.9

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

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

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

2b3.9. Results of Risk Stratification Analysis:

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

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

_____

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

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

2018 Submission

To demonstrate meaningful differences in performance, NCQA calculates an inter-quartile range (IQR) for each indicator. 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, NCQA calculates an independent sample t-test of the performance difference between two randomly selected plans at the 25th and 75th percentile. The t-test method calculates a testing statistic based on the sample size, performance rate, and standardized error of each plan. 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 plans' performance is significantly different from each other. Using this method, we compared the performance rates of two randomly selected plans, one plan in the 25th percentile and another plan in the 75th percentile of performance. We used these two plans as examples of measures entities. However, the method can be used for comparison of any two measured entities

2012 submission

Comparison of means and percentiles; analysis of variance against established benchmarks: if sample size is >400, we would use an analysis of variance.

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

2018 Submission

Variation in Performance across Health Plans for Initiation Indicator in 2017 Data

	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Commercial	1,617	36.7	7.7	29.4	33.0	35.9	39.2	42.0	6.2	0.0250
Medicare	1,328	34.4	13.1	15.2	27.0	35.1	41.7	48.3	14.7	<0.001
Medicaid	3,967	42.3	7.4	33.7	38.6	42.2	46.4	50.2	7.8	<0.001

EP: Eligible Population, the average denominator size across plans submitting to HEDIS

IQR: Interquartile range

P-value: P-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile. P-values are less than 0.05.

Variation in Performance across Health Plans for Engagement Indicator in 2017 Data

	Avg. EP	Avg.	SD	10th	25th	50th	75th	90th	IQR	p-value
Commercial	1,617	13.4	4.1	8.7	11.0	13.3	15.9	18.3	4.9	<0.001
Medicare	1,328	4.2	2.9	0.8	2.2	3.7	5.6	8.1	3.4	.0064
Medicaid	3,963	13.6	5.9	6.1	9.1	13.7	17.7	21.4	8.6	<0.001

EP: Eligible Population, the average denominator size across plans submitting to HEDIS

IQR: Interquartile range

P-value: P-value of independent samples t-test comparing plans at the 25th percentile to plans at the 75th percentile. P-values are less than 0.05.

2012 submission

Initiation:

Medicaid

Measurement Year:	2009	2010	2011
AVE:	44.52	44.35	42.93
N:	61	68	79
Min:	17.74	22.72	23.86
Max:	69.09	76.71	78.88
SD:	10.01	10.31	10.96
P10:	32.74	31.78	30
P25	37.21	38.42	35.68
P50	43.79	43.92	40.81
P75	51.26	48.79	48.84
P90	57.33	57.31	60.72

Medicare

Measurement Year:	2009	2010	2011
AVE:	46.55	48.89	48.02
N:	268	306	368
Min:	5.19	12.12	7.32
Max:	84.85	98.23	98.24
SD:	13.94	15.72	17.07
P10:	29.25	27.41	27.31
P25:	38.48	38.88	36.17
P50:	46.83	49.17	46.08
P75:	54.7	56.9	57.62
P90:	64.29	70.27	74.11

Commercial

Measurement Year:	2009	2010	2011
AVE:	42.46	42.28	41.89
N:	415	402	392
Min:	14.71	12.9	16.67

Measurement Year:	2009	2010	2011
Max:	70.18	72.65	69.77
SD:	7.4	7.32	7.51
P10:	33.47	34.03	33.01
P25:	38.6	38.2	37.42
P50:	42.2	41.79	41.81
P75:	46.67	46.27	45.71
P90:	51.33	50.6	50.27

Engagement:

Medicaid

Measurement Year:	2009	2010	2011
AVE:	12.43	12.31	14.19
N:	61	68	79
Min:	0	0.99	0.5
Max:	55.57	54.26	41.44
SD:	11.45	10.73	9.79
P10:	1.69	2.34	2.02
P25:	3.46	4.15	5.72
P50:	10.06	10.18	14.53
P75:	16.79	17.6	20.52
P90:	21.7	21.42	25.89

Medicare

Measurement Year:	2009	2010	2011
AVE:	5.36	4.51	4.02
N:	268	311	366
Min:	0	0	0
Max:	41.79	35.64	26.25
SD:	6.23	4.17	3.46
P10:	0.7	0.8	0.56
P25:	1.97	2.08	1.71
P50:	3.13	3.52	3.19
P75:	6.32	5.78	5.61
P90:	11.63	8.53	7.95

Commercial

Measurement Year:	2009	2010	2011
AVE:	16.2	15.93	15.78
N:	415	402	392
Min:	0	1.61	0.85
Max:	53.4	46.99	46.45
SD:	5.7	5.88	5.6
P10:	9.74	8.51	9.54
P25:	12.43	12.19	12.01
P50:	15.85	15.61	15.56

Measurement Year:	2009	2010	2011
P75:	19.82	19.19	18.68
P90:	22.46	22.19	22.09

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

2018 Submission

The results above indicate there is a 6.2-14.7% gap in performance for the initiation indicator, and a 3.4-8.6% gap in performance for the engagement indicator between plans performing at the 25th and 75th percentiles. The difference between plan performance at the 25th and 75th percentile is statistically significant for both indicator rates across all product lines.

In commercial plans, there is a 6.2 percentage point gap between 25th and 75th percentile plans for the initiation of treatment indicator rate. This gap represents an average 97 more patients who have initiated treatment for alcohol or other drug abuse or dependence within 14 days of their new diagnosis in high performing plans compared to low performing plans (estimated from average health plan eligible population). For the engagement in treatment indicator rate, there is a 4.9 percentage point gap between 25th and 75th percentile commercial plans. This gap represents an average 79 more patients who have engaged in treatment within the 34 days following initiation of treatment for alcohol or other drug abuse or dependence in high performing plans compared to low performing plans (estimated from average health plan eligible population).

In Medicare plans, there is a 14.7 percentage point gap between 25th and 75th percentile plans for the initiation of treatment indicator rate. This gap represents an average 195 more patients who have initiated treatment for alcohol or other drug abuse or dependence within 14 days of their new diagnosis in high performing plans compared to low performing plans (estimated from average health plan eligible population). For the engagement in treatment indicator rate, there is a 3.4 percentage point gap between 25th and 75th percentile Medicare plans. This gap represents an average 45 more patients who have engaged in treatment within the 34 days following initiation of treatment for alcohol or other drug abuse or dependence in high performing plans compared to low performing plans (estimated from average health plan eligible population).

In Medicaid plans, there is a 7.8 percentage point gap between 25th and 75th percentile plans for the initiation of treatment indicator rate. This gap represents an average 309 more patients who have initiated treatment for alcohol or other drug abuse or dependence within 14 days of their new diagnosis in high performing plans compared to low performing plans (estimated from average health plan eligible population). For the engagement in treatment indicator rate, there is an 8.6 percentage point gap between 25th and 75th percentile Medicaid plans. This gap represents an average 341 more patients who have engaged in treatment within the 34 days following initiation of treatment for alcohol or other drug abuse or dependence in high performing plans compared to low performing plans (estimated from average health plan eligible population).

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

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

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

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

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

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

_____

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

2018 Submission

HEDIS measures apply to enrolled members in a health plan, and NCQA has a rigorous audit process to ensure the eligible population and numerator events for each measure are correctly identified and reported. The audit process is designed to verify primary data sources used to populate measures and ensure specifications are correctly implemented.

The HEDIS Compliance Audit addresses the following functions, as applicable:

- Information practices and control procedures
- Sampling methods and procedures
- Data integrity
- Compliance with HEDIS specifications
- Analytic file production
- Reporting and documentation

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

2018 Submission

HEDIS addresses missing data in a structured way through its audit process. HEDIS measures apply to enrolled members in a health plan, and NCQA-certified auditors use standard audit methodologies to assess whether data sources are missing data. If a data source is found to be missing data, and the issues cannot be rectified, the auditor will assign a "materially biased" designation to the measure for that reporting plan, and the rate will not be used in any analyses. Once measures, new or re-evaluated, are added to HEDIS, NCQA conducts an analysis to assess the measure's feasibility for implementation in the field. This analysis includes an assessment of how many plans report valid rates vs. rates that are materially biased (or have other issues, such as small denominators). These considerations are weighed in the deliberation process before measures are approved for public reporting.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the*

selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data)

2018 Submission

Plans collect this measure using all administrative data sources. NCQA's audit process checks that plans' measure calculations are not biased due to missing data.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in a combination of electronic sources

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

N/A

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

Attachment:

3c. Data Collection Strategy

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

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

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

NCQA recognizes that, despite the clear specifications defined for HEDIS measures, data collection and calculation methods may vary, and other errors may taint the results, diminishing usefulness of HEDIS data for managed care organization (MCO) comparison. In order for HEDIS to reach its full potential, NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO's ability to comply with HEDIS specifications. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.

The HEDIS Compliance Audit addresses the following functions:

1) Information practices and control procedures
2) Sampling methods and procedures
3) Data integrity
4) Compliance with HEDIS specifications
5) Analytic file production
6) Reporting and documentation

In addition to the HEDIS audit, NCQA provides a system to allow "real-time" feedback from measure users. Our Policy Clarification Support System receives thousands of inquiries each year on over 100 measures. Through this system, NCQA responds immediately to questions and identifies possible errors or inconsistencies in the implementation of the measure. This system is vital to the regular re-evaluation of NCQA measures.

Input from NCQA auditing and the Policy Clarification Support System informs the annual updating of all HEDIS measures including updating value sets and clarifying the specifications. Measures are re-evaluated on a periodic basis and when there is a significant change in evidence. During re-evaluation information from NCQA auditing and Policy Clarification Support System is used to inform evaluation of the scientific soundness and feasibility of the measure.

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

Broad public use and dissemination of these measures, without modification, are encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Modifications to, and/or commercial use of, a measure requires the prior written consent of NCQA and is subject to a license at the discretion of NCQA. As used herein, "commercial use" refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure .

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
	Public Reporting Health Plan Rating https://www.ncqa.org/hedis/reports-and-research/ratings-methodology-and-guidelines/ Annual State of Health Care Quality http://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality Payment Program Medicaid Adult Core Set https://www.medicaid.gov/medicaid/quality-of-care/downloads/medicaid-adult-core-set-manual.pdf Merit Based Incentive Payment System (MIPS) https://qpp.cms.gov/mips/quality-measures Centers for Medicare & Medicaid Services (CMS) Health Insurance Marketplace Quality Initiatives: Health Insurance Exchange Quality Rating System (QRS) http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Health-Insurance-Marketplace-Quality-Initiatives.html Regulatory and Accreditation Programs NCQA Accreditation http://www.ncqa.org/tabid/123/Default.aspx Quality Improvement (external benchmarking to organizations) Quality Compass http://www.ncqa.org/hedis-quality-measurement/quality-measurement-products/quality-compass Annual State of Health Care Quality http://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality

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

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

MEDICAID ADULT CORE SET: There are a core set of health quality measures for Medicaid-enrolled adults. The Medicaid Adult Core Set was identified by the Centers of Medicare & Medicaid (CMS). The data collected from these measures helps CMS to better understand the quality of health care that adults enrolled in Medicaid receive nationally. Beginning in January 2014 and annually thereafter, the Secretary is required to publicly report the information that states voluntarily report to CMS on the quality of health care received by adults enrolled in Medicaid.

MERIT BASED INCENTIVE PAYMENT SYSTEM (MIPS) QUALITY PAYMENT PROGRAM (QPP): Eligible clinicians who elect to participate in MIPs earn a performance-based payment adjustment to Medicaid payments upon submission of evidence which attests that they provided high quality, efficient care supported by technology.

Eligible clinicians can select up to six quality measures to report to CMS, including one outcome measure, that best fit their needs or specialty. The data collected from this program will help CMS to better understand the quality of health care that Medicare enrollees receive nationally.

HEALTH INSURANCE EXCHANGE QUALITY RATING SYSTEM (QRS): Qualified Health Plan (QHP) issuers and Multi-State Plan (MSP) issuers that offered coverage through a Health Insurance Marketplace (Marketplace) in the year prior to the current year are required to collect and submit QRS measure data to CMS. CMS produces quality ratings on a 5-star scale for each issuer in each State. Health plan level clinical quality measures and survey measures based on questions from the Qualified Health Plan Enrollee Experience Survey (QHP Enrollee Survey) are included in the QRS measure set. CMS collects data and calculates quality ratings for each QHP issuer's product type within each state and applies these ratings to each product type's QHPs in that State.

STATE OF HEALTH CARE ANNUAL REPORT: This measure is publicly reported nationally and by geographic regions in the NCQA State of Health Care annual report. This annual report published by NCQA summarizes findings on quality of care. In 2017, the report included results from calendar year 2016 for health plans covering a record 182 million people, or 43 percent of the U.S. population

HEALTH PLAN RATINGS/REPORT CARDS: This measure is used in the calculation of health plan ratings, which are reported on the NCQA website annually. These ratings are based on a plan's performance on their HEDIS, CAHPS and accreditation standards scores. In 2017, a total of 521 Medicare Advantage health plans, 614 commercial health plans and 294 Medicaid health plans across 50 states, D.C., Guam, Puerto Rico, and the Virgin Islands were included in the Ratings.

HEALTH PLAN ACCREDITATION: This measure is used in scoring for accreditation of Medicare Advantage Health Plans. As of Fall 2017, a total of 184 Medicare Advantage health plans were scored for accredition using this measure among others covering 9.2 million Medicare beneficiaries; 451 commercial health plans covering 113 million lives; and 125 Medicaid health plans covering 35 million lives. Health plans are scored based on performance compared to national benchmarks.

QUALITY COMPASS: This measure is used in Quality Compass which is an indispensable tool used for selecting health plans, conducting competitor analysis, examining quality improvement and benchmarking plan performance. Provided in this tool is the ability to generate custom reports by selecting plans, measures, and benchmarks (averages and percentiles) for up to three trended years. Results in table and graph formats offer comparison of plans' performance against competitors or benchmarks.

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

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

N/A

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

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

Health plans that report HEDIS calculate their rates and know their performance when submitting to NCQA. NCQA publicly reports rates across all plans and also creates benchmarks in order to help plans understand how they perform relative to other plans. Public reporting and benchmarking are effective quality improvement methods.

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

NCQA publishes HEDIS results annually in our Quality Compass tool. NCQA also presents data at various conferences and webinars. For example, at the annual HEDIS Update and Best Practices Conference, NCQA presents results from all new measures' first year of implementation or analyses from measures that have changed significantly. NCQA also regularly provides technical assistance on measures through its Policy Clarification Support System.

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

Describe how feedback was obtained.

NCQA measures are evaluated regularly. During this "reevaluation" process, we seek broad input on the measure, including input on performance and implementation experience. We use several methods to obtain input, including vetting of the measure with several multi-stakeholder advisory panels, public comment posting, and review of questions submitted to the Policy Clarification Support System. This information enables NCQA to comprehensively assess a measure's adherence to the HEDIS Desirable Attributes of Relevance, Scientific Soundness and Feasibility.

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

In general, health plans have not reported significant barriers to implementing this measure, as it uses the administrative data collection method. Questions have generally centered around minor clarification of the specifications, such as calculating days of medication treatment and questions about the supporting guidelines for the measure. NCQA responded to all questions to ensure consistent implementation of the specifications.

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

This measure has been deemed a priority measure by NCQA and other entities, as illustrated by its use in programs such as the CMS Quality Rating System (QRS), CMS Merit-Based Incentive Payment System (MIPS) Program, and the Medicaid Adult Core Set.

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

Feedback has not required modification to this measure.

Improvement

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

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

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

Over the past three years, this measure has shown slight improvement across health plans (see section 1b.2 for summary of data from health plans), although overall, still demonstrating that there is a significant gap in care and room for improvement across all product lines.

Overall, mean performance and distribution for the initiation of treatment (initiation indicator) was relatively similar among the Medicare, Medicaid and commercial products. Starting in 2018, data was stratified by diagnosis cohort (i.e., alcohol, opioid, or other drug abuse and dependence) to understand with more

granularity how different subpopulations were initiating and engaging in treatment. For the initiation indicator, higher performance was seen among members with a diagnosis of opioid abuse and dependence than members with diagnoses of alcohol or other drug abuse and dependence.

Mean performance and distribution for the engagement in ongoing treatment (engagement indicator) among Medicaid and commercial products were very similar. However, performance was about 10 percentage points lower than what is observed in the Medicaid and commercial products. For the engagement indicator, higher performance was again seen among members with a diagnosis of opioid abuse and dependence than members with diagnoses of alcohol or other drug abuse and dependence.

4b2. Unintended Consequences

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

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

There were no identified unexpected benefits during implementation of this measure

NCQA recognizes that, despite the clear specifications defined for HEDIS measures, data collection and calculation methods may vary, and other errors may taint the results, diminishing the usefulness of HEDIS data for managed care organization (MCO) comparison. In order for HEDIS to reach its full potential, NCQA conducts an independent audit of all HEDIS collection and reporting processes, as well as an audit of the data which are manipulated by those processes, in order to verify that HEDIS specifications are met. NCQA has developed a precise, standardized methodology for verifying the integrity of HEDIS collection and calculation processes through a two-part program consisting of an overall information systems capabilities assessment followed by an evaluation of the MCO´s ability to comply with HEDIS specifications (. NCQA-certified auditors using standard audit methodologies will help enable purchasers to make more reliable "apples-to-apples" comparisons between health plans.

The HEDIS Compliance Audit addresses the following functions:

1) information practices and control procedures
2) sampling methods and procedures
3) data integrity
4) compliance with HEDIS specifications
5) analytic file production
6) reporting and documentation

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

N/A

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

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

5b. Competing Measures

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

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

N/A

Appendix

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

No appendix Attachment:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): National Committee for Quality Assurance

Co.2 Point of Contact: Bob, Rehm, nqf@ncqa.org, 202-955-1728-

Co.3 Measure Developer if different from Measure Steward: National Committee for Quality Assurance

Co.4 Point of Contact: Kristen, Swift, swift@ncqa.org, 202-955-5174-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

NCQA Behavioral Health Measurement Advisory Panel

Katharine Bradley, MD, MPH, Senior Investigator, Kaiser Permanente Washington Health Research Institute

Christopher Dennis, MD, MBA, FAPA, Chief Behavioral Health Officer, Landmark Health

Ben Druss, MD, MPH, Professor, Emory University

Frank Ghinassi, PhD, ABPP, President & CEO, Rutgers University Behavioral Health Care

Connie Horgan, ScD, Professor and Director, Institute for Behavioral Health, Brandeis University

Laura Jacobus-Kantor, PhD, Chief, Quality, Evaluation and Performance Branch, SAMHSA, HHS

Jeffrey Meyerhoff, MD, National Medical Director for Medicare and Retirement, Optum Behavioral Health

Harold Pincus, MD, Professor and Vice Chair--Department of Psychiatry, College of Physicians and Surgeons

Co-Director, Irving Institute for Clinical and Translational Research --Columbia University

Director of Quality and Outcomes Research--New York –Presbyterian Hospital

Senior Scientist--RAND Corporation

Michael Schoenbaum, PhD, Senior Advisor for Mental Health Services, Epidemiology and Economics, National Institute of Mental Health

John Straus, MD, Medical Director Special Projects, Beacon Health Options

NCQA Committee on Performance Measurement (CPM)

Andrew Baskin, MD, National Medical Director, Quality & Provider Performance Measurement, Aetna

Helen Darling, MA. Strategic Advisor on Health Benefits & Health Care

Andrea Gelzer, MD, MS, FACP, Senior Vice President & Corporate Chief Medical Officer, AmeriHealth Caritas

Kate Goodrich, MD, MHS, Director, Center for Clinical Standards and Quality and CMS Chief Medical Officer, Centers for Medicare and Medicaid Services

David Grossman, MD, MPH. Senior Associate Medical Director, Market Strategy and Public Policy. Washington Permanente Medical Group

Christine S. Hunter, MD, Chief Medical Officer, US Office of Personnel Management

Jeffrey Kelman, MD, MMSc., Chief Medical Officer, Center for Medicare, United States Department of Health and Human Services (DHHS)

Nancy Lane, Ph.D., Independent Consultant

Bernadette Loftus, MD, Associate Executive Director for the Mid-Atlantic States, The Permanente Medical Group

Adrienne Mims, MD, MPH, Vice President, Chief Medical Officer, Alliant Quality

Amanda Parsons, MD, MBA, Vice President, Community & Population Health, Montefiore Health System

Wayne Rawlins, MD, MBA, Chief Medical Officer, ConnectiCare

Rodolfo Saenz, MD, MMM, FACOG, Obstetrician/Gynecologist, Medical Director of Quality, Riverside Medical Clinic

Eric C. Schneider, MD, MSc, FACP, Senior Vice President, Policy and Research, The Commonwealth Fund

Marcus Thygeson, MD, MPH, Chief Health Officer, Bind Benefits

JoAnn Volk, MA, Research Professor, Georgetown University Center on Health Insurance Reforms

Lina Walker, PhD, Vice President of Health Security, AARP

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2004

Ad.3 Month and Year of most recent revision: 01, 2012

Ad.4 What is your frequency for review/update of this measure? As needed, based on feedback from the field and changes to clinical guidelines and evidence.

Ad.5 When is the next scheduled review/update for this measure? 12, 2019

Ad.6 Copyright statement: The HEDIS® measures and specifications were developed by and are owned by the National Committee for Quality Assurance (NCQA). The HEDIS measures and specifications are not clinical guidelines and do not establish a standard of medical care. NCQA makes no representations, warranties, or endorsement about the quality of any organization or physician that uses or reports performance measures and NCQA has no liability to anyone who relies on such measures or specifications. NCQA holds a copyright in these materials and can rescind or alter these materials at any time. These materials may not be modified by anyone other than NCQA. Anyone desiring to use or reproduce the materials without modification for a non-commercial purpose may do so without obtaining any approval from NCQA. All commercial uses must be approved by NCQA and are subject to a license at the discretion of NCQA. ©2018 NCQA, all rights reserved.

Calculated measure results, based on unadjusted HEDIS specifications, may not be termed "Health Plan HEDIS rates" until they are audited and designated reportable by an NCQA-Certified Auditor. Such unaudited results should be referred to as "Unaudited Health Plan HEDIS Rates." Accordingly, "Heath Plan HEDIS rate" refers to and assumes a result from an unadjusted HEDIS specification that has been audited by an NCQA-Certified HEDIS Auditor.

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

Content reproduced with permission from HEDIS, Volume 2: Technical Specifications for Health Plans. To purchase copies of this publication, including the full measures and specifications, contact NCQA Customer Support at 888-275-7585 or visit www.ncqa.org/publications.

Ad.7 Disclaimers: This HEDIS® performance measure is not a clinical guideline and does not establish a standard of medical care and has not been tested for all potential applications.

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

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

Ad.8 Additional Information/Comments: NCQA Notice of Use. Broad public use and dissemination of these measures, without modification, are encouraged and NCQA has agreed with NQF that noncommercial uses do not require the consent of the measure developer. Modifications to, and/or commercial use of, a measure requires the prior written consent of NCQA and is subject to a license at the discretion of NCQA. As used herein, "commercial use" refers to any sale, license, or distribution of a measure for commercial gain, or incorporation of a measure into any product or service that is sold, licensed, or distributed for commercial gain, even if there is no actual charge for inclusion of the measure.