NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

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

Brief Measure Information

NQF #: 3589

Corresponding Measures:

De.2. Measure Title: Prescription or administration of pharmacotherapy to treat opioid use disorder (OUD)

Co.1.1. Measure Steward: RTI International

De.3. Brief Description of Measure: This measure reports the percentage of a provider's patients who were Medicaid beneficiaries ages 18 to 64 with an OUD diagnosis who filled a prescription for, or were administered or ordered, a FDA-approved medication to treat OUD within 30 days of the first attributable OUD treatment encounter with that provider.

1b.1. Developer Rationale: In 2019, opioids were involved in 49,912 overdose deaths in the United States, a 6.6% annual increase from 2018. The two clinical milestones likely to have the greatest impact on lowering the risk of mortality from opioid use disorders are initiation of medications to treat opioid use disorder (MOUD) initiation and retention on MOUD (Ball & Ross, 1991; Degenhardt et al., 2011; Sordo et al., 2017; Volkow et al., 2019). Randomized clinical trials and observational studies find that individuals with an opioid use disorder (OUD) who are treated with an FDA-approved opioid use disorder medication (i.e., methadone, buprenorphine, naltrexone) have better outcomes than individuals who do not receive OUD medications such as larger reductions in mortality, opioid usage and relapses, rates of infectious disease, and emergency department and inpatient admissions, as well as improved functioning in major life domains (Clark et al., 2014; Clark et al., 2015; D´Onofrio et al., 2015; Fullerton et al., 2014; Haley et al., 2019; Jarvis et al., 2018; Larochelle et al., 2018; Lo-Ciganic et al., 2016; Ma et al., 2019; Mark et al., 2020; Mattick et al., 2003; Mattick et al., 2009; Mattick et al., 2014; Minozzi et al., 2011; Parran et al., 2010; Pierce et al., 2016; Schwarz et al., 2012; Sordo et al., 2017; Syed & Keating, 2013; Thomas et al., 2014; Williams et al., 2020; Woody et al., 2014). As described below, comprehensive reviews conclude that the evidence for the effectiveness of MOUD relative to non-medication based OUD treatment is of high to moderate quality (American Society of Addiction Medicine, 2015, 2020; Center for Substance Abuse Treatment, 2004, Center for Substance Abuse Treatment, 2005; Fullerton et al., 2014; Thomas et al., 2014, Mattick et al., 2014, Mattick et al., 2009, SAMHSA, 2020).

NQF endorsed Use of Pharmacotherapy (NQF #3400) – a measure of the percentage of Medicaid beneficiaries ages 18 to 64 with an OUD who filled a prescription for or were administered or ordered an FDA-approved

medication for opioid use disorder (MOUD) during the measure year - at the health plan and Medicaid program level. The implementation of this measure is hypothesized to lead to more people receiving MOUD, improved health care outcomes, higher treatment costs, but lower overall costs because high-cost acute care utilization is reduced and the comorbidities associated with OUD are treated more effectively (See Figure 1. Logic model) (Busch et al., 2017; Clark et al., 2014; Clark et al., 2015; Florence et al., 2013; Mohlman et al., 2016; Mark et al., 2020; National Institute of Drug Abuse, 2019; Nielsen et al., 2016; Ronquest et al., 2018). The potential untended consequences of the measure may include diversion and accidental overdoses, such as by children. A recent review of 17 studies finds that most people use illicit buprenorphine to self-medicate (to manage opioid withdrawal symptoms or achieve or maintain abstinence from other opioids), while a smaller percentage use it to get high (Chilcoat et al., 2019). Between 2004 and 2011, there were 5,222 emergency department visits by children ages 1 to 5 involving accidental ingestion of buprenorphine (Crane, 2017). To put that number in perspective, there were 8.2 million prescriptions for buprenorphine in 2012. Buprenorphine is available in an extended-release injectable form with no diversion potential and no potential for accidental ingestion.

Despite the clear benefits of MOUD, many practitioners do not offer them to their patients with OUD. In 2018, only 40% of specialty addiction treatment facilities offered medications to treat opioid use disorder (Mark et al., 2020). In 2015, between 31% and 37% of patients with OUD in specialty facilities received medications for OUD (SAMHSA, 2017). Many regions of the country lack an adequate supply of buprenorphine waived professionals (Abraham et al., 2020; Andrilla et al., 2020). A recent study in Massachusetts reported that only 30% of those who survived an opioid overdose received medications for OUD in the year after their overdose (Larochelle, 2018).

A 2014 American Society of Addiction Medicine expert panel recommended that the Use of Pharmacotherapy Measure (NQF #3400) be created as a measure at the clinician level (ASAM, 2014). They note that individual providers could use the measure for quality improvement and to monitor their practices (ASAM, 2014). The ASAM expert panel noted that the measure would be useful even though there is no set time when a patient with OUD should be initiated on MOUD or an absolute benchmark level to determine the appropriate percentages of patients on MOUD. They explain that ensuring timely treatment with OUD medications is important as data show that individuals who receive MOUD are more likely to be retained in addiction treatment (Mattick et al., 2014, Timko et al., 2016). Further, improved outcomes have been demonstrated for patients who initiate buprenorphine to treat OUD in emergency departments (D'Onofrio et al., 2015). Finally, as noted above, a large and robust portfolio of research highlights the significant mortality and morbidity benefits of MOUD.

A measure of use of MOUD for OUD is being reported and used at the provider level in four state Medicaid programs to help identify providers who could benefit from technical assistance (New York, Massachusetts, West Virginia, Delaware). The measure is also being used at the provider-level in Centers for Medicare and Medicaid (CMS) behavioral health home demonstrations; however, it has not been endorsed at the provider level (CMS, 2019).

S.4. Numerator Statement: Beneficiaries ages 18 to 64 with an OUD who filled a prescription for, or were administered or ordered, an FDA-approved medication for the treatment of OUD within 30 days of the first attributable encounter with an OUD diagnosis with the provider.

S.6. Denominator Statement: Number of Medicaid ages 18 – 64 beneficiaries with at least one medical claim for an encounter with an OUD diagnosis with that provider (where the provider is identified by a National Provider Identifier (NPI) code).

S.8. Denominator Exclusions: Dual eligible Medicare/Medicaid beneficiaries are excluded. Rationale: Individuals who are covered under Medicare would receive coverage for follow up treatment medications (e.g. medication

assisted treatment) under Medicare Part D and Medicare Part D claims are not captured in Medicaid claims databases. Therefore, follow-up would be missed.

Individuals under 18 are excluded. Rationale: There is limited evidence regarding the efficacy of MOUD for this population.

Individuals over 64 are excluded: Rationale: Most individuals over age 64 are covered under Medicare. Services covered by Medicare would not be capture in the Medicaid claims data and therefore follow-up treatment would be missed.

De.1. Measure Type: Process

S.17. Data Source: Claims, Enrollment Data

S.20. Level of Analysis: Clinician: Individual, Facility

Preliminary Analysis: New Measure

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

The developer provides the following evidence for this measure:

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

Evidence Summary

- This new claims-based measure at the clinician and facility level reports the percentage of a provider's patients who were Medicaid beneficiaries ages 18 to 64 with an OUD diagnosis who filled a prescription for, or were administered or ordered, an FDA-approved medication to treat OUD within 30 days of the first attributable OUD treatment encounter with that provider.
- Developer provides a logic model depicting the relationship between structural changes to improve medication for opioid use disorder (MOUD), processes to improve MOUD and the resulting patient outcomes.

- Developer cites recommendations from the 2020 American Society of Addiction Medicine (ASAM) National Practice Guideline for the Use of Medications in the Treatment of Addiction Involving Opioid Use:
    - p. 11, "Methadone is a recommended treatment for patients with opioid use disorder, who are able to give informed consent and have no specific contraindication for this treatment."
    - p.12, "Buprenorphine is a recommended treatment for patients with opioid use disorder, who are able to give informed consent and have no specific contraindication for this treatment."
    - p. 13, "Extended-release injectable naltrexone is a recommended treatment for preventing relapse to opioid use disorder in patients who are no longer physically dependent on opioids, able to give informed consent, and have no contraindications for this treatment."
    - The guidelines do not provide grades.
- Developer provides additional systematic reviews supporting the use of methadone and buprenorphine which conclude that:
    - "Overall, there is a high level of evidence for the effectiveness of [Methadone Maintenance Treatment [MMT)] in improving treatment retention and decreasing illicit opioid use", noting a "large number of trials" suggesting overall evidence rating for MMT is high.
    - "[Buprenorphine Maintenance Treatment] (BMT)" is associated with improved outcomes compared with placebo for individuals and pregnant women with opioid use disorders", noting that the grade assigned to the evidence was high. The author's stated that "because of the large number of trials, the overall evidence for BMT was rated as high. Thus, the level of research evidence is similar for BMT and MMT."
- Developer cites Cocharane collaboration review and meta-analysis of 31 trials of buprenorphine vs methadone or placebo, noting high to moderate quality of evidence range.
- Developer cites meta-analysis of 11 studies of methadone vs no opioid replacement therapy, noting good evidence for positive outcomes and effective management of OUD with methadone.
- Developer cites a 2005 Substance Abuse and Mental Health Service Administration (SAMHSA) publication from the Center for Substance Abuse Treatment which notes that "Research supports the perspective that opioid addiction is a medical disorder that can be treated effectively with medications when they are administered under conditions consistent with their pharmacological efficacy and when treatment includes necessary supportive services such as psychosocial counseling, treatment for co-occurring disorders, medical services, and vocational rehabilitation." This is an ungraded recommendation.
- Developer cites a 2020 SAMHSA Treatment Improvement Protocol for Medications for Opioid Use Disorder, which concludes that:
    - p. 1-5, "The TIP expert panel strongly recommends informing all patients with OUD about the risks and benefits of treatment of OUD with all FDA-approved medications. Alternatives to these treatments and their risks and benefits should be discussed. Patients should receive access to such medications if clinically appropriate and desired by the patients."
    - p. 1-8, "Patients can take medication for OUD on a short-term or long-term basis. However, patients who discontinue OUD medication generally return to illicit opioid use."
    - p. 1-8, "The best results occur when a patient receives medication for as long as it provides a benefit. This approach is often called 'maintenance treatment.'"
    - These statements were not graded.

Questions for the Committee:

- Is the existing evidence sufficient to warrant a national performance measure?

Guidance from the Evidence Algorithm

(Box 1) Process measure (Box 3) Systematic reviews graded (Box 4) QQC provided (Box 5) High quality, quantity and consistency of evidence HIGH

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.

- Summary data for individual clinicians and facilities was provided:

Summary Data of Observed Scores

n	Mean	SD	Min	10th	25th	50th	75th	90th	Max
Individual Clinicians									
5344	44%	35%	0%	0%	10%	37%	79%	93%	100%
Hospitals/Facilities/Agencies									
4054	31%	28%	0%	3%	9%	22%	45%	80%	100%

- Mean performance for clinicians was 44%; mean performance for facilities was 31%

Disparities

- Developer provides a table that indicates disparities in the use of medications to treat opioid use disorder.
- The table describes the percentage of patients diagnosed with an opioid use disorder who received medications to treat opioid use disorder by gender and race.
  - Men are less likely to receive medications than women.
  - Blacks are less likely to receive medications than Whites.
  - Hispanics/Latinos are more likely to receive medications than Whites or Blacks.

Questions for the Committee:

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

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

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

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures–are you aware of any new

studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- excellent evidence base
- Evidence is sufficient.
- The evidence relates directly to the outcome being measured and is aligned with ASAM treatment guidelines.
- Evidence is high
- strong evidence
- The evidence relates directly to the measure. The developer includes evidence regarding current prescription rates for MOUD for patients with OUD, as well as information regarding the risks of not prescribing MOUD. They also include clinical practice guidelines. Yes, the evidence is sufficient.
- Evidence exists, but not sure it warrants a national measure.
- While I understand that medications are effective for OUD, I am at a loss why individual clinicians are held accountable rather than systems. Most PCPs are not trained in using these medications. And since there is inclusion of rural health clinics, FQHCs, etc., they would fall under this measure. Accountability at a facility or system level: yes. Individual level, absolutely not. Did I miss something?
- Evidence applies directly to the process measure. The evidence supports the benefits of initiation and retention of MOUD in OUD patients. The process measure has the potential to improve appropriate use of FDA-approved MOUD, which ultimately could reduce OUD mortality, overdoses, comorbidities, and poor overall functioning. I am not aware of any new information that changes the evidence base for this measure that has not been cited in the submission.
- Empirical data cited by submission applies directly to the measure and desired outcomes (e.g. prescription for OUD). OUD is a growing problem in the United States and literature cites evidence to support use of methadone, buprenorphine, and naltrexone to treat OUD.
- Evidence seems sufficient to warrant a national performance measure.
- Process
- There is evidence that use of MOUD improves outcomes and that timely initiation of MOUD improves participation and engagement in treatment. Theoretically the use of sample fall within the parameter and would be included in the results but practically I know of no system that able to capture such actions into their results. At heart this is a process measure which contributes to the real outcomes of lower morbidity and mortality and improved functioning, health, employment and quality of life. As I see how it is being used and discussed at the health-plan level I see that many folks are starting to consider the prescribing or administering the appropriate med as and end unto itself
- The evidence is robust and strong to warrant a national performance measure.

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

- yes. demonstrated, including disparity data.
- Sufficiently demonstrates performance gaps.

- The submission includes extensive gaps with respect to gender and ethnicity.
- Performance gap info for SUD providers only. The gap is evident and disparities information was provided. However; what about the performance gap among other provider types like MH and Primary care.
- significant gap: clinician 44% (SD=35%), facility: 31% (SD=28%)
- Yes, performance data was provided for individual clinicians and facilities. A minority of clinicians and facilities provide prescriptions for MOUS. Given the national crisis of opioid use and overdoses, this is an important issue that warrants a national performance measure. Data was provided for population subgroups demonstrating that men receive less prescriptions than women, Hispanics/Latinos receive more prescriptions than Whites, who receive more prescriptions than Blacks. These are disparities in care that can significantly impact health outcomes, recovery, function and life expectancy.
- Performance gaps exist.
- yes, a gap. Some disparities, but would like to see more around SES, rurality etc.
- There is high variability across individual clinicians as well as treatment facilities warranting a national performance measure that can help improve these gaps in care. Disparities were noted, specifically men are less likely to receive medications than women. Blacks are less likely to receive medications than Whites. Hispanics/Latinos are more likely to receive medications than Whites or Blacks.
- There is a high gap in care, as the mean score for clinicians was 44% and the mean facility score was 31%, both indicating room for improvement in pharmacotherapy to treat OUD.
- Does seem to be sizable variability in early applications, suggesting major room for improvement. In terms of disparities, the fact that men are less likely to receive medications than women and blacks are less likely to receive medications than whites suggests it IS important to address disparities as performance focus.
- yes there are opportunities of improvement with men, blacks, Hispanics Latinos.
- There clearly is a performance gap and data to show racial and gender disparities.
- Yes. There is a clear gap in care and notable disparities.

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.

Composite measures only:

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

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

Measure was evaluated by NQF Staff

NQF Staff Evaluation Summary:

Reliability

- Developer conducted several score-level reliability analyses, including:
  - Parametric ANOVA with effect size calculations
    - Clinicians: $F = 71.17$; $\eta 2 = 0.58$; $\omega 2 = 0.57$
    - Facilities: $F = 84.84$; $\eta 2 = 0.44$; $\omega 2 = 0.44$
  - Intra-unit reliability (IUR) – 0.99 for clinicians and facilities
  - Beta-binomial signal to noise analysis:

Measures	n	Mean	SD	Min	10th	25th	50th	75th	90th	Max
Individual Clinicians	5344	0.95	0.05	0.83	0.88	0.92	0.96	0.99	1	1
Hospitals/Facilities/Agencies	4054	0.95	0.05	0.76	0.86	0.92	0.97	0.99	1	1

- The reliability testing results suggest that the measure is reliable.
  - The F-statistic for the signal-to-noise ratio indicates that the measure scores are significantly different.
  - Additional analysis indicates that the providers in the sample can be appropriately ranked on performance of the measure.

Validity

- Convergent validity is established by empirically showing that measures that are conceptually related are statistically correlated to one another.
- Pearson product moment correlation coefficients were calculated between the measure and:
  - The Continuity of Care After Inpatient or Residential SUD Treatment at the Provider Level (the "SUD Follow-Up" measure): $r = 0.39$

    - Hospitalization or ED visit associated with SUD or overdose within 30 days after an encounter with the provider: r = 0.39
- Convergent validity testing was conducted using a common method for conceptually similar measures, producing results that were statistically significant, directionally appropriate and of moderate strength.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

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

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

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

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

- High
- No concerns.
- The submission includes results from reliability testing and has preliminary rating of high.
- The definition of provider is unclear. Does this measure apply to both BH providers and primary care providers? Why not include in the denominator providers who prescribed MAT during the second and third visit? Why exclude providers with less than 10 attributed clients?
- clear specifications
- Data elements are clearly defined. Codes with descriptors are provided and the steps are clear. I do not have concerns that this measure can be consistently implemented as reliability estimates are high.
- Measure is reliable.
- OK
- Data elements are clearly defined, with adequate sampling, and good reliability demonstrated through a standard approach. I do not have concerns about the likelihood that this measure can be consistently implemented.
- I have no concerns about implementation as the data used to measure are claims data and is straightforward

- Measure specifications seem adequate.
- No concerns
- Reliability seems adequate
- These are provided and clear. No concerns about it being implemented.

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

- No
- No.
- No
- Highly reliable
- No. in addition to signal to noise, calculating effect size was a plus. Pearson r's were .39 when examining convergent validity
- I do not.
- No concerns.
- not really
- No because ANOVA, signal-to-noise, IUR, Adam's rho for individual clinicians and hospitals/facilities/agencies were all were within acceptable range.
- No
- Score-level reliability analyses seem adequate.
- No concerns.
- No
- I do not have concerns with the reliability testing provided.

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

- no
- No.
- No
- Moderate
- no
- I do not.
- No concerns.
- only moderate at best
- I do not.
- No
- Empirical validity tests seem adequate. No concerns.
- No concerns.
- Nothing significant
- No concerns for the validity testing provided.

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

- none identified
- Acceptable.
- I concur with the exclusion of dual eligible beneficiaries as their medications are paid for by Part D. I see no inappropriate exclusions.
- no concerns
- given data source, unable to stratify by most social determinants of health
- All exclusions are in alignment with the evidence and no groups are excluded inappropriately. The developer does not include information on risk adjustment as it is not applicable since it looks at claims for typical care.
- No concerns.
- meh
- No concerns about the patients excluded from the measure. Risk adjustment - n/a.
- N/A
- Exclusions are appropriate. Risk adjustment or stratification N/A.
- No risk adjustment in this measure.
- NA
- While I understand the reasoning for not including patients 65+ or dual eligible, this is a growing group of patients who need MOUD but are often neglected.

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

- none
- I am satisfied with the testing of the measure's validity.
- I concur with the NQF staff finding on validity testing. I do not see a threat to validity.
- No concerns
- no concerns
- This measure includes data on prescriptions for OUD using claims data. Analyses of this measure indicate that there are meaningful differences in quality at the facility and provider level and among

differences in population subgroups. Yes, measure is specified precisely indicating comparable results. The developer rare instances of missing data, such that it should not have an impact on the measure.

- No concerns.
- I am not convinced that differences at the clinician level accurately reflect QOC.
- The F statistic indicates a statistically significant difference in performance between providers and Eta-squared and omega-squared values indicate large effect size. Comparability - n/a. Missing data - reported to be rare and does not impact the measure..
- No
- No concerns. Missing data not expected to be an issue. Analysis addresses meaningful differences. Comparability N/A.
- No threat to validity.
- I do not see a lot of evidence specifying that prescribing or administering the med within the timeframe of exactly 30 DAYS is evidence based per se. I do believe that it's a reasonable marker of timeliness.
- The main issue I have related to the definition of OUD diagnoses used by the developers. Identifying OUD in claims data is difficult and there is poor specificity of ICD codes, particularly for "opioid dependence" which could include people on long-term opioid therapy who have physiologic dependence, but not OUD. Therefore this measure could include people without OUD (who would then not be treated with MOUD).

Criterion 3. Feasibility

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

3. **Feasibility** is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
    - Measure uses claims, where data elements are routinely generated and used during care delivery. Coded by someone other than person obtaining original information. All data elements are in defined fields in a combination of electronic sources.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- high. already demonstrated feasible in NY

- Feasible.
- I concur with the NQF staff finding of high feasibility.
- The measure is feasible-uses claims data
- feasible using claims data
- Some data are generated during usual care, other data are coded after delivery such as diagnosis. All data are available in electronic form. I do not have concerns about how the data can be put into use.
- There are feasibility concerns, is medication available - cost, prescribers, adherence. Not clear if this measure is feasible.
- feasible.
- No concerns regarding feasibility. Data elements are generated or collected by healthcare personnel during the provision of care and all data elements are in defined fields in electronic claims.
- No concerns about feasibility since the measure uses claims data that are already routinely collected
- No concerns about feasibility – data elements are all part of regular claims. New York State's extensive vetting with providers, patients, and other stakeholders lends support for its feasibility.
- No concerns .
- It is feasible
- No concerns.

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details

- New York Office of Addiction Supports and Services, Shatterproof ATLAS, a related measure is being used as part of the CMS Medicaid Adult and Home Core Sets Program.
- Geographic area and number and percentage of accountable entities and patients included: New York state (approximately 274 addiction treatment facilities), Shatterproof ATLAS (approximately 400 addiction treatment providers across 4 states, New York, Massachusetts, Delaware, and West Virginia).
- Level of measurement and setting: Provider of addiction treatment.

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

Feedback on the measure by those being measured or others

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

Additional Feedback: N/A

Questions for the Committee:

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

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

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

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

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

Improvement results

- Measure has not been implemented

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

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

Potential harms: None identified.

Additional Feedback: N/A

***Questions for the Committee**:*

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

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

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

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

- not reported?
- Yes
- I cannot discern from the submission whether or not the measure is being publicly reported.
- yes.
- not publicly reported, but used in 4 Medicaid programs, integrated into NY O Addiction qi
- The measure has not yet been implemented and does not have results to report. Data will be presented in a portal for those who have access. There is a credible plan for quality improvement through providing results on prescription for MOUD. It is currently being used for external benchmarking for facilities and internal to specific organizations. It aims to improve prescriptions for OUD and quality improvement plans describe how to change physician behaviors. Quality improvement plans should also include recommendations for managing the underlying causes of opioid use, for example, pain management. Such recommendations might include services such as occupational therapy, physical therapy, pain psychology, health behavior changes.
- New measure.
- Some data from shatterproof ATLAS
- Measure is not publicly reported but is currently being used and will be more broadly used in accountability programs. Feedback was garnered in the development phase of this measure from state medicaid programs, commercial health plans, addiction treatment providers, patients, families, and other experts.
- N/A, as this measure is not in use; however developer states that measure was developed using feedback from state Medicaid plans and others using similar measures and during an NQF session to gather feedback.
- No concerns. New York State's extensive vetting with providers, patients, and other stakeholders lends support for its use. Having a measure at the provider/clinician level is important to OUD treatment improvement.
- Yes

- The MA/health plan level measure is already being used and this one is likely to follow suit.
- No concerns.

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

- yes. appears useful for quality and performance
- Measure is usable, however where and/or how are limits on prescribers and/or lack of prescribers accounted for?
- Given the underutilization of MAT for OUD this measure can be used to improve quality of care for Medicaid beneficiaries. Benefit far outweighs any potential harm.
- benefits outweigh the harms
- no concerns
- If prescriptions for OUD do improve health and functional outcomes and reduce risks, this measure can be used to assist physicians and facilities in understanding their current behavior with writing prescriptions and engage in quality improvement efforts to improve. Yes, there is a clear rationale for how performance results could be used to provide high quality, efficient healthcare. See above for additional considerations for providing education to physicians. Since the measure relies on claims data for services already rendered and will provide feedback to providers and facilities that could support quality improvement, there are no unintended consequences of reviewing the data. Supporting increased medication for OUD does come with some risks, as noted by the developer. There are risks associated with children taking prescribed medications, but the likelihood of this is low given the potential benefits.
- Unintended consequences of patients diverting medications and selling them. Children of patients may has access to the medications.
- yes.
- The performance results can be used to further the goal of high-quality, efficient healthcare by identifying and improving quality/process gaps (inadequate training, clinician staffing, ineffective workflows, etc.) which may be contributing to the low rates of MOUD prescribing for patients with OUD who are clinically appropriate for and desire medication assisted treatment. I believe the potential benefits (namely decreased mortality) outweigh the potential harms in this measure.
- N/A, not yet used in an accountability program. No harms identified.
- No concerns.
- Benefits of the medications for oud outweigh any potential harms.
- It should help outcomes
- Especially with the removal of the x-waiver, there should not be barriers for patient with OUD accessing MOUD. MOUD need to be made available for patients and covered by Medicaid. The hope is this measure would encourage payers and policymakers to continue to move towards equitable access for MOUD.

Criterion 5: Related and Competing Measures

Related or competing measures

- 3175: Continuity of Pharmacotherapy for Opioid Use Disorder
- 3400: Use of Pharmacotherapy for Opioid Use Disorder (OUD)

Harmonization

Developer notes that the measure is harmonized with NQF3400: Use of Pharmacotherapy for Opioid Use Disorder (OUD). The same OUD code and pharmacotherapy codes are included in both. The difference between NQF 3400 and this measure (Prescription or administration of pharmacotherapy to treat OUD), is that this measure is meant to be used at the provider level.

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

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

- Yes. this is more specific with 30 day interval. There is considerable evidence of MOUD continuity at 30-60 days predicts 6 month follow up. I prefer current measure.
- The submission notes that the measure is harmonized with NQF3400: Use of Pharmacotherapy for Opioid Use Disorder (OUD).
- harmonized
- harmonized with NQF 3400
- Two other measures are listed by the developer and an explanation for how one has been harmonized is included. This measure differs because it provides data at the provider level.
- no.
- I think developing measures that clearly are the same or similar except for level of analysis is duplicative (NCQA). This creates measure bloat.
- 3175 and 3400, harmonized with the latter.
- Measure has been harmonized with NQF3400, using same coding; however current measure reports results at provider level while NQF3400 reports at plan level. I believe both are useful.
- While there are related measures, they do not appear to be competing in any way, and measure specifications (for NQF3400: Use of Pharmacotherapy for Opioid Use Disorder) are adequately harmonized.
- None
- NQF 3400 uses the same definitions as this to perform an identical analysis but at the MA/health plan level. I think that our NQF BH Measurement Committee should ask the developer to analyze the results of both measures to see what can be learned by comparing and contrasting the results. Historically in my region most providers were in the networks of almost every health plan and the plans results were generally indistinguishable. As more consolidation occurred and managed care products became a higher percentage, more narrow networks became prominent one could see more differentiation.

Those systems where a plan and delivery system largely shared responsibility and incentives to improve care for the same population have had improved outcomes. If others agree that this might be useful, we could reserve some time at a future meeting for this.

- Yes, but I agree the focus on provider level is a must.

Public and Member Comments

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

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

NQF Staff Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: NQF 3589

Measure Title: Prescription or administration of pharmacotherapy to treat opioid use disorder (OUD)

Type of measure:

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

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

Data Source:

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

Level of Analysis:

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

   Submission document: Items S.1-S.22

2. **Briefly summarize any concerns about the measure specifications.**
   - No concerns identified by staff

RELIABILITY: TESTING

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

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?
   ☐ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**
   Submission document: Testing attachment, section 2a2.2
   - Developer conducted several score-level reliability analyses, including:
     - Parametric ANOVA with effect size calculations
       - Clinicians: F = 71.17; η2 = 0.58; ω2 = 0.57
       - Facilities: F = 84.84; η2 = 0.44; ω2 = 0.44
     - Intra-unit reliability (IUR) – 0.99 for clinicians and facilities
     - Beta-binomial signal to noise analysis:

Measures	n	Mean	SD	Min	10th	25th	50th	75th	90th	Max
Individual Clinicians	5344	0.95	0.05	0.83	0.88	0.92	0.96	0.99	1	1
Hospitals/Facilities/Agencies	4054	0.95	0.05	0.76	0.86	0.92	0.97	0.99	1	1

7. **Assess the results of reliability testing**
   Staff concurs with developer assessment of the results: The reliability testing results suggest that the measure is highly reliable. The F-statistic for the signal-to-noise ratio indicates that the measure scores are significantly different while subsequent reliability statistics (the IUR and Adam's rho) indicate a large effect size, empirically substantiating that the measure can discern underlying performance between providers.
8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.
   Submission document: Testing attachment, section 2a2.2
   ☒ **Yes**
   ☐ **No**
   ☐ **Not applicable** (score-level testing was not performed)
9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?
   Submission document: Testing attachment, section 2a2.2
   ☐ **Yes**
   ☐ **No**

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

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

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

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

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

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

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

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

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    - None identified by staff

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

    Submission document: Testing attachment, section 2b4.

    - None identified by staff

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

    - None identified by staff

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

    Submission document: Testing attachment, section 2b6.

    - None identified by staff

16. **Risk Adjustment**

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

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

    ☐ Yes ☐ No ☒ Not applicable

    16c. **Social risk adjustment:**

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

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

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

VALIDITY: TESTING

17. **Validity testing level: ☒ Measure score ☐ Data element ☐ Both**
18. **Method of establishing validity of the measure score:**
    - ☐ **Face validity**
    - ☒ **Empirical validity testing of the measure score**
    - ☐ **N/A (score-level testing not conducted)**
19. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    - Convergent validity is established by empirically showing that measures that are conceptually related are statistically correlated to one another.
    - Pearson product moment correlation coefficients were calculated between the measure and:
        - The Continuity of Care After Inpatient or Residential SUD Treatment at the Provider Level (the "SUD Follow-Up" measure): r = 0.39
        - Hospitalization or ED visit associated with SUD or overdose within 30 days after an encounter with the provider: r = 0.39
20. **Assess the results(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.3

    - Convergent validity testing was conducted using a common method for conceptually similar measures, producing results that were statistically significant, directionally appropriate and of moderate strength.
21. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    Submission document: Testing attachment, section 2b1.

    - ☒ **Yes**
    - ☐ **No**
    - ☐ **Not applicable** (score-level testing was not performed)
22. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

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

    - ☐ **Yes**
    - ☐ **No**
    - ☒ **Not applicable** (data element testing was not performed)
23. **OVERALL RATING OF VALIDITY considering the results and scope of all testing and analysis of potential threats.**

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

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

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

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

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

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

ADDITIONAL RECOMMENDATIONS

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

- None identified.

Developer Submission

NQF #: 3589

Corresponding Measures:

De.2. Measure Title: Prescription or administration of pharmacotherapy to treat opioid use disorder (OUD)

Co.1.1. Measure Steward: RTI International

De.3. Brief Description of Measure: This measure reports the percentage of a provider's patients who were Medicaid beneficiaries ages 18 to 64 with an OUD diagnosis who filled a prescription for, or were administered or ordered, a FDA-approved medication to treat OUD within 30 days of the first attributable OUD treatment encounter with that provider.

1b.1. Developer Rationale: In 2019, opioids were involved in 49,912 overdose deaths in the United States, a 6.6% annual increase from 2018. The two clinical milestones likely to have the greatest impact on lowering the risk of mortality from opioid use disorders are initiation of medications to treat opioid use disorder (MOUD) initiation and retention on MOUD (Ball & Ross, 1991; Degenhardt et al., 2011; Sordo et al., 2017; Volkow et al., 2019). Randomized clinical trials and observational studies find that individuals with an opioid use disorder (OUD) who are treated with an FDA-approved opioid use disorder medication (i.e., methadone, buprenorphine, naltrexone) have better outcomes than individuals who do not receive OUD medications such as larger reductions in mortality, opioid usage and relapses, rates of infectious disease, and emergency department and inpatient admissions, as well as improved functioning in major life domains (Clark et al., 2014; Clark et al., 2015; D´Onofrio et al., 2015; Fullerton et al., 2014; Haley et al., 2019; Jarvis et al., 2018; Larochelle et al., 2018; Lo-Ciganic et al., 2016; Ma et al., 2019; Mark et al., 2020; Mattick et al., 2003; Mattick et al., 2009; Mattick et al., 2014; Minozzi et al., 2011; Parran et al., 2010; Pierce et al., 2016; Schwarz et al., 2012; Sordo et al., 2017; Syed & Keating, 2013; Thomas et al., 2014; Williams et al., 2020; Woody et al., 2014). As described below, comprehensive reviews conclude that the evidence for the effectiveness of MOUD relative to non-medication based OUD treatment is of high to moderate quality (American Society of Addiction Medicine, 2015, 2020; Center for Substance Abuse Treatment, 2004, Center for Substance Abuse Treatment, 2005; Fullerton et al., 2014; Thomas et al., 2014, Mattick et al., 2014, Mattick et al., 2009, SAMHSA, 2020).

NQF endorsed Use of Pharmacotherapy (NQF #3400) – a measure of the percentage of Medicaid beneficiaries ages 18 to 64 with an OUD who filled a prescription for or were administered or ordered an FDA-approved medication for opioid use disorder (MOUD) during the measure year - at the health plan and Medicaid program level. The implementation of this measure is hypothesized to lead to more people receiving MOUD, improved health care outcomes, higher treatment costs, but lower overall costs because high-cost acute care utilization is reduced and the comorbidities associated with OUD are treated more effectively (See Figure 1. Logic model) (Busch et al., 2017; Clark et al., 2014; Clark et al., 2015; Florence et al., 2013; Mohlman et al., 2016; Mark et al., 2020; National Institute of Drug Abuse, 2019; Nielsen et al., 2016; Ronquest et al., 2018). The potential untended consequences of the measure may include diversion and accidental overdoses, such as by children. A recent review of 17 studies finds that most people use illicit buprenorphine to self-medicate (to manage opioid withdrawal symptoms or achieve or maintain abstinence from other opioids), while a smaller percentage use it to get high (Chilcoat et al., 2019). Between 2004 and 2011, there were 5,222 emergency department visits by children ages 1 to 5 involving accidental ingestion of buprenorphine (Crane, 2017). To put that number in perspective, there were 8.2 million prescriptions for buprenorphine in 2012. Buprenorphine is available in an extended-release injectable form with no diversion potential and no potential for accidental ingestion.

Despite the clear benefits of MOUD, many practitioners do not offer them to their patients with OUD. In 2018, only 40% of specialty addiction treatment facilities offered medications to treat opioid use disorder (Mark et al.,

2020). In 2015, between 31% and 37% of patients with OUD in specialty facilities received medications for OUD (SAMHSA, 2017). Many regions of the country lack an adequate supply of buprenorphine waived professionals (Abraham et al., 2020; Andrilla et al., 2020). A recent study in Massachusetts reported that only 30% of those who survived an opioid overdose received medications for OUD in the year after their overdose (Larochelle, 2018).

A 2014 American Society of Addiction Medicine expert panel recommended that the Use of Pharmacotherapy Measure (NQF #3400) be created as a measure at the clinician level (ASAM, 2014). They note that individual providers could use the measure for quality improvement and to monitor their practices (ASAM, 2014). The ASAM expert panel noted that the measure would be useful even though there is no set time when a patient with OUD should be initiated on MOUD or an absolute benchmark level to determine the appropriate percentages of patients on MOUD. They explain that ensuring timely treatment with OUD medications is important as data show that individuals who receive MOUD are more likely to be retained in addiction treatment (Mattick et al., 2014, Timko et al., 2016). Further, improved outcomes have been demonstrated for patients who initiate buprenorphine to treat OUD in emergency departments (D'Onofrio et al., 2015). Finally, as noted above, a large and robust portfolio of research highlights the significant mortality and morbidity benefits of MOUD.

A measure of use of MOUD for OUD is being reported and used at the provider level in four state Medicaid programs to help identify providers who could benefit from technical assistance (New York, Massachusetts, West Virginia, Delaware). The measure is also being used at the provider-level in Centers for Medicare and Medicaid (CMS) behavioral health home demonstrations; however, it has not been endorsed at the provider level (CMS, 2019).

S.4. Numerator Statement: Beneficiaries ages 18 to 64 with an OUD who filled a prescription for, or were administered or ordered, an FDA-approved medication for the treatment of OUD within 30 days of the first attributable encounter with an OUD diagnosis with the provider.

S.6. Denominator Statement: Number of Medicaid ages 18 – 64 beneficiaries with at least one medical claim for an encounter with an OUD diagnosis with that provider (where the provider is identified by a National Provider Identifier (NPI) code).

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

Individuals under 18 are excluded. Rationale: There is limited evidence regarding the efficacy of MOUD for this population.

Individuals over 64 are excluded: Rationale: Most individuals over age 64 are covered under Medicare. Services covered by Medicare would not be capture in the Medicaid claims data and therefore follow-up treatment would be missed.

De.1. Measure Type: Process

S.17. Data Source: Claims, Enrollment Data

S.20. Level of Analysis: Clinician: Individual, Facility

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

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

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

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

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

MOUD_Evidence_11_19_2020.docx

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

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

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM — Evidence (subcriterion 1a)

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

Measure Title: Prescription or administration of pharmacotherapy to treat opioid use disorder (OUD)

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

Date of Submission: October 29, 2020

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

☐ Outcome: Click here to name the health outcome

☐ Patient-reported outcome (PRO):

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

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

☐ Process: **Use of pharmacotherapy for opioid use disorder (OUD) at the provider level**

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

☐ Composite:

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

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

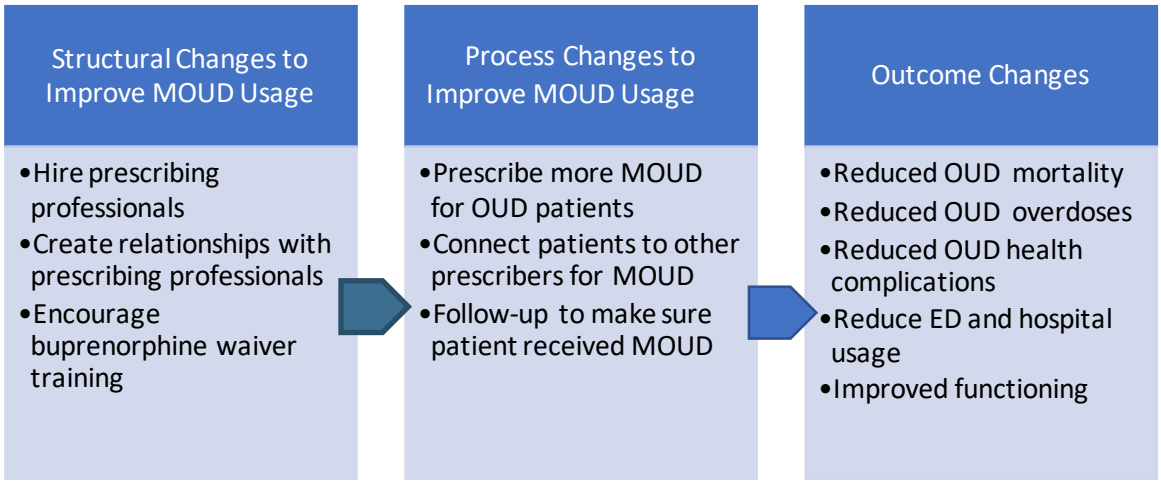
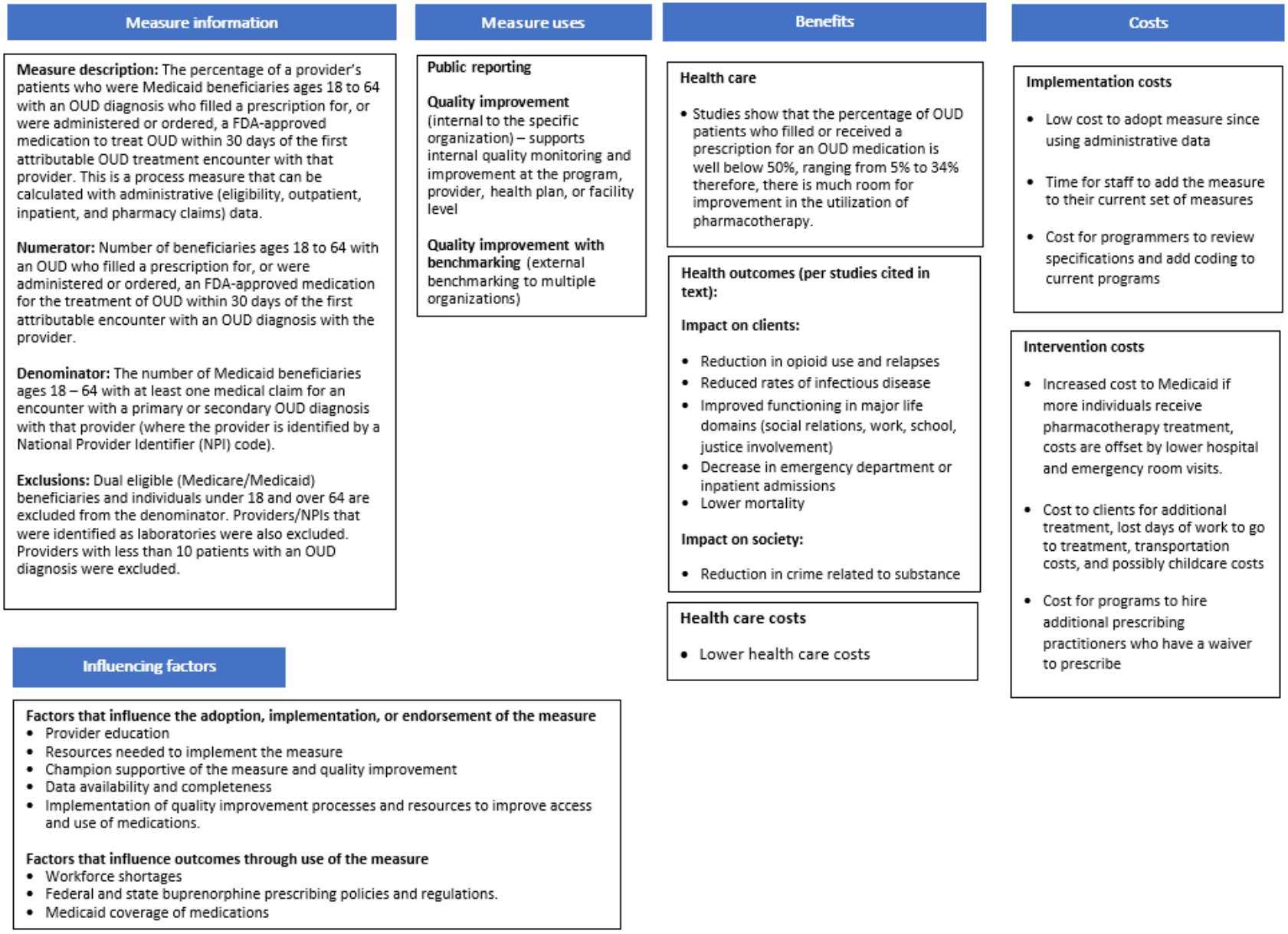


Figure 2. Alternate Model for Prescription or administration of pharmacotherapy to treat opioid use disorder (OUD)



In 2019, opioids were involved in 49,912 overdose deaths in the United States, a 6.6% annual increase from 2018. The two clinical milestones likely to have the greatest impact on lowering risk of mortality from opioid use disorders are initiation of medications to treat opioid use disorder (MOUD) initiation and retention on MOUD (Ball & Ross, 1991; Degenhardt et al., 2011; Sordo et al., 2017; Volkow et al., 2019). Randomized clinical trials and observational studies find that individuals with an opioid use disorder (OUD) who are treated with an FDA-approved opioid use disorder medication (i.e., methadone, buprenorphine, naltrexone) have better outcomes than individuals who do not receive OUD medications including larger reductions in mortality, opioid usage and relapses, rates of infectious disease, and emergency department and inpatient admissions, as well as improved functioning in major life domains (Clark et al., 2014; Clark et al., 2015; D'Onofrio et al., 2015; Fullerton et al., 2014; Haley et al., 2019; Jarvis et al., 2018; Larochelle et al., 2018; Lo-Ciganic et al., 2016; Ma et al., 2019; Mark et al., 2020; Mattick et al., 2003; Mattick et al., 2009; Mattick et al., 2014; Minozzi et al., 2011; Parran et al., 2010; Pierce et al., 2016; Schwarz et al., 2012; Sordo et al., 2017; Syed & Keating, 2013; Thomas et al., 2014; Williams et al., 2020; Woody et al., 2014). As described below, comprehensive reviews conclude that the evidence for the effectiveness of MOUD relative to non-medication based OUD treatment is of high to moderate quality (American Society of Addiction Medicine, 2015, 2020; Center for Substance Abuse Treatment, 2004, Center for Substance Abuse Treatment, 2005; Fullerton et al., 2014; Thomas et al., 2014, Mattick et al., 2014, Mattick et al., 2009, SAMHSA, 2020).

NQF endorsed *Use of Pharmacotherapy* (NQF #3400) – a measure of percentage of Medicaid beneficiaries ages 18 to 64 with an OUD who filled a prescription for or were administered or ordered an FDA-approved medication for opioid use disorder (MOUD) during the measure year - at the health plan and Medicaid program level. The implementation of this measure is hypothesized to lead to more people receiving MOUD, improved health care outcomes, higher treatment costs, but lower overall costs because high cost acute care utilization are reduced and the comorbidities associated with OUD are treated more effectively (See Figure 1. Logic model) (Busch et al., 2017; Clark et al., 2014; Clark et al., 2015; Florence et al., 2013; Mohlman et al., 2016; Mark et al., 2020; National Institute of Drug Abuse, 2019; Nielsen et al., 2016; Ronquest et al., 2018). The potential untended consequences of the measure may include diversion and accidental overdoses such as by children. A recent review of 17 studies finds that most people use illicit buprenorphine to self-medicate (to manage opioid withdrawal symptoms or achieve or maintain abstinence from other opioids), while a smaller percentage use it to get high (Chilcoat et al., 2019). Between 2004 and 2011 there were 5,222 emergency department visits by children ages 1 to 5 involving accidental ingestion of buprenorphine (Crane, 2017). To put that number in perspective, there were 8.2 million prescriptions for buprenorphine in 2012. Buprenorphine is available in an extended release injectable form which has no diversion potential and no potential for accidental ingestion.

Despite the clear benefits of MOUD, many practitioners do not offer them to their patients with OUD. In 2018, only 40% of specialty addiction treatment facilities offered medications to treat opioid use disorder (Mark et al., 2020). In 2015, between 31% and 37% of patients with OUD in specialty facilities received medications for OUD (SAMHSA, 2017). Many regions of the country lack adequate supply of buprenorphine waived professionals (Abraham et al., 2020, Andrilla et al., 2020). A recent study in Massachusetts reported that only 30% of those who survived an opioid overdose received medications for OUD in the year after their overdose (Larochelle, 2018).

A 2014 American Society of Addiction Medicine expert panel recommended that the Use *of Pharmacotherapy Measure* (NQF #3400) be created as a measure at the clinician level (ASAM, 2014). They note that individual providers could use the measure for quality improvement and to monitor their practices (ASAM, 2014). The ASAM expert panel noted that the measure would be useful despite the fact

that there is no set time in which a patient with OUD should be initiated on MOUD or an absolute benchmark level to determine the appropriate percentages of patients on MOUD. They explain that ensuring timely treatment with OUD medications is important as data show that individuals who receive MOUD are more likely to be retained in addiction treatment (Mattick et al., 2014, Timko et al., 2016). Further, improved outcomes have been demonstrated for patients who initiate buprenorphine to treat OUD in emergency departments (D'Onofrio et al., 2015). Finally, as noted above, a large and robust portfolio of research highlights the significant mortality and morbidity benefits of MOUD.

A measure of use of MOUD for OUD is being reported and used at the provider level is being used in four state Medicaid programs to help identify providers who could benefit from technical assistance (New York, Massachusetts, West Virginia, Delaware). The measure is also being used at the provider-level in Centers for Medicare and Medicaid (CMS) behavioral health home demonstrations, however, it has not been endorsed at the provider level (CMS, 2019).

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

Systematic Review	Evidence
Source of Systematic Review 1: • Title • Author • Date • Citation, including page number • URL	Title: The ASAM National Practice Guideline for the Treatment of Opioid Use Disorder – 2020 Focused Update Author: American Society of Addiction Medicine Date: Adopted by the ASAM Board of Directors December 18, 2019 Citation: American Society of Addiction Medicine. (2020). National practice guideline for the use of medications in the treatment of addiction involving opioid use. URL: https://www.asam.org/docs/default-source/quality-science/npg-jam-supplement.pdf?sfvrsn=a00a52c2_2
Quote the guideline or recommendation verbatim about the process, structure or	p. 11, "Methadone is a recommended treatment for patients with opioid use disorder, who are able to give informed consent and have no specific

Systematic Review	Evidence
intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	contraindication for this treatment." p.12, "Buprenorphine is a recommended treatment for patients with opioid use disorder, who are able to give informed consent and have no specific contraindication for this treatment." p. 13, "Extended-release injectable naltrexone is a recommended treatment for preventing relapse to opioid use disorder in patients who are no longer physically dependent on opioids, able to give informed consent, and have no contraindications for this treatment."
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Not Applicable. The guidelines cited above do not provide grades (e.g., USPSTF grades A, B, etc.).
Provide all other grades and definitions from the evidence grading system	The ASAM practice guidelines were developed by using the RAND appropriateness method. Guideline developers reviewed existing literature and guidelines, treatment scenarios, appropriateness ratings, and other documents. An independent panel convened by ASAM oversaw development of the guidelines.
Grade assigned to the **recommendation** with definition of the grade	Not Applicable. The guidelines cited above do not provide grades (e.g., USPSTF grades A, B, etc.).
Provide all other grades and definitions from the recommendation grading system	The ASAM practice guidelines were developed by using the RAND appropriateness method. Guideline developers reviewed existing literature and guidelines, treatment scenarios, appropriateness ratings, and other documents. An independent panel convened by ASAM oversaw development of the guidelines.
Body of evidence: Quantity – how many studies? Quality – what type of studies?	The number of studies referenced in support of the guidelines was 215. The quality of the evidence selected indicates that included studies were considered to be of high quality by an expert panel.
Estimates of benefit and consistency across studies	Buprenorphine pharmacotherapy is recommended for the treatment of opioid use disorder (American Society of Addiction Medicine, 2020). Buprenorphine relieves drug cravings without producing euphoria, and with reduced risk of dangerous and adverse effects compared with full agonist opioids. In addition to its pharmacological properties, an advantage of buprenorphine is that it can be prescribed in office-based treatment settings. The FDA approved buprenorphine in 2002, making it the first medication eligible to be prescribed by certified physicians through the Drug Addiction Treatment Act of 2000 (DATA 2000).105 Through DATA 2000, physicians may apply for waivers to prescribe certain

Systematic Review	Evidence
	narcotic schedule III, IV, or V medications, including buprenorphine, from their office settings. This provision of the act expands access to community-based treatment options and mitigates the need to receive treatment through more specialized, and often less available, OTPs. However, buprenorphine may also be administered in an OTP setting with similar program and administration requirements to those for methadone. Recent legislation has further expanded the types of practitioners who can prescribe buprenorphine for the treatment of opioid use disorder. The Comprehensive Addiction and Recovery Act (CARA) signed into law in July 2016 extended the authority to prescribe buprenorphine to qualifying NPs and PAs who obtain a waiver. The SUPPORT for Patients and Community Act (Congress.gov) signed into law in October 2018 further expanded buprenorphine prescribing privileges (through October 1, 2023) to qualifying clinical nurse specialists, certified registered nurse anesthetists, and certified nurse midwives. Methadone, a slow-acting opioid agonist, is an effective treatment for opioid withdrawal management and the treatment of opioid use disorder. Methadone is taken orally so that it reaches the brain slowly, dampening the rewarding effect that can occur with other routes of administration while preventing withdrawal symptoms. Methadone has been used since the 1960 s to treat heroin addiction and remains an effective treatment option. Many studies have demonstrated its superiority to medication-free approaches. In the United States, Methadone is only available through approved OTPs, where it is dispensed to patients on a daily or almost daily basis in the initial stages of treatment, and in acute care settings (under limited circumstances). Federal and state laws allow take-home doses for patients who have demonstrated treatment progress and are judged to be at low risk for diversion, and for whom the therapeutic benefits of take-home doses outweigh the risks. Extended-release injectable naltrexone and under limited circumstances, oral naltrexone, are effective treatments recommended for patients to prevent relapse to opioid use disorder, are able to give informed consent, are fully withdrawn from opioids, and have no specific contraindications for this treatment.
What harms were identified?	Sufficient evidence points to the safety and efficacy of buprenorphine for the treatment of OUD (Parran et al., 2010). The risk of fatal overdose on buprenorphine is substantially lower than that associated with the use of other opioid medications such as methadone because of the ceiling effects of buprenorphine across a wide range of doses (American Society of Addiction Medicine, 2015).

Systematic Review	Evidence
Source of Systematic Review 2: • Title • Author • Date • Citation, including page number • URL	Title: Medication-Assisted Treatment With Methadone: Assessing the Evidence Author: Fullerton, C. A., Kim, M., Thomas, C. P., Lyman, D. R., Montejano, L. B., Dougherty, R. H., Daniels, A. S., Ghose, S. S., & Delphin-Rittmon, M. E. Date: 2014 Citation: Fullerton, C. A., Kim, M., Thomas, C. P., Lyman, D. R., Montejano, L. B., Dougherty, R. H., ... & Delphin-Rittmon, M. E. (2014). Medication-assisted treatment with methadone: assessing the evidence. *Psychiatric Services*, *65*(2), 146-157. URL: N/A
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.	"Overall, there is a high level of evidence for the effectiveness of [Methadone Maintenance Treatment [MMT)] in improving treatment retention and decreasing illicit opioid use (see box on previous page). Research findings regarding the impact of MMT on many secondary outcomes, such as mortality, drug-related HIV risk behaviors, and criminal activity, are less conclusive but suggest positive trends. Finally, research has not conclusively shown positive impacts on sex-related HIV risk behaviors, nonopioid illicit drug or alcohol use, or other social consequences. Methadone maintenance doses above 60 mg confer greater efficacy in retention and suppression of illicit opioid use; however, there is limited evidence that doses above 100 mg provide additional benefits. No evidence has emerged to delineate the duration of MMT beyond an indefinite period. Although MMT generally is believed to reduce mortality risk among individuals with opioid dependence, methadone is also associated with significant adverse events, such as respiratory depression and cardiac arrhythmias, in the presence of rapid titrations or other risk factors. There is no clear evidence that structured psychotherapy provided in addition to the psychosocial support normally offered at methadone treatment centers conveys additional benefit. MMT improves pregnancy-related outcomes by reducing illicit drug use and increasing treatment retention. However, newborn infants of mothers treated with methadone during pregnancy may be born with NAS irrespective of the methadone dose used by the mothers." "MMT is an important treatment option for opioid dependence. Providers, consumers, and family members should be educated about the benefits of MMT in helping individuals manage opioid use disorders and about appropriate ways to avoid the significant adverse events that can occur with methadone. Providers and consumers need to be educated regarding appropriate doses to improve efficacy and appropriate initiation to minimize adverse events. Because of MMT's relative efficacy, efforts should be made to increase access to MMT for all individuals who struggle with opioid use disorders. Directors of state mental health and substance abuse agencies and community health organizations should look for methods to increase access to MMT, and purchasers of health care services should cover appropriately monitored MMT."

Systematic Review	Evidence
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	"Because of the large number of trials included as individual studies or as part of review articles, the overall evidence rating for MMT is high. Several meta-analyses, reviews, and RCTs representing more than three independent RCTs have reported on the primary outcomes of MMT, which are retention in treatment and reduction of illicit opioid use. In addition, meta-analyses, reviews, RCTs, and quasi-experimental studies representing more than three RCTs or two RCTs and two quasi-experimental studies have addressed secondary outcomes such as other illicit drug use, HIV risk behaviors, criminal behaviors, heroin craving, and mortality." "Evidence for the effectiveness of methadone maintenance treatment: high. Evidence clearly shows that MMT has a positive impact on: retention in treatment, illicit opioid use. Evidence is less clear but suggestive that MMT has a positive impact on: mortality, illicit drug use (nonopioid), drug-related HIV risk behaviors, criminal activity, evidence suggests that MMT has little impact on: sex-related HIV risk behaviors."
Provide all other grades and definitions from the evidence grading system	"Three levels of evidence (high, moderate, and low) were used to indicate the overall research quality of the collection of studies. Ratings were based on predefined benchmarks that considered the number of studies and their methodological quality. If ratings were dissimilar, a consensus opinion was reached." "High ratings indicate confidence in the reported outcomes and are based on three or more RCTs with adequate designs or two RCTs plus two quasiexperimental studies with adequate designs. Moderate ratings indicate that there is some adequate research to judge the service, although it is possible that future research could influence reported results. Moderate ratings are based on the following three options: two or more quasiexperimental studies with adequate design; one quasi-experimental study plus one RCT with adequate design; or at least two RCTs with some methodological weaknesses or at least three quasi-experimental studies with some methodological weaknesses. Low ratings indicate that research for this service is not adequate to draw evidence based conclusions. Low ratings indicate that studies have nonexperimental designs, there are no RCTs, or there is no more than one adequately designed quasi-experimental study."
Grade assigned to the **recommendation** with definition of the grade	*
Provide all other grades and definitions from the recommendation grading system	*
Body of evidence:	"Authors reviewed meta-analyses, systematic reviews, and individual studies of MMT from 1995 through 2012. Databases searched were

Systematic Review	Evidence
• Quantity – how many studies? • Quality – what type of studies?	PubMed, PsycINFO, Applied Social Sciences Index and Abstracts, Sociological Abstracts, Social Services Abstracts, and Published International Literature on Traumatic Stress. The literature search found 7 RCTs and two retrospective, quasi-experimental studies. 15 reviews or meta-analyses that examined multiple studies [were also included].
Estimates of benefit and consistency across studies	"Research supports MMT's positive impact on treatment retention and suppression of heroin use, particularly at higher methadone doses. Findings regarding secondary outcomes are mixed, although there is general support that MMT has a positive impact on criminal activity associated with heroin use, as well as on mortality and risk behaviors for HIV and hepatitis C infection." "In general, these and later studies found that when MMT is provided at adequate dose levels, it is more effective than no medication treatment in retaining patients in treatment and reducing illicit opioid use." "MMT during pregnancy was associated with decreased illicit opioid use, increased rates of prenatal retention in treatment, decreased pregnancy complications, and generally improved fetal outcomes."
What harms were identified?	"MMT has been found to put newborn infants at risk for neonatal abstinence syndrome (NAS)—a condition characterized by dysfunction of the autonomic nervous system, gastrointestinal tract, and respiratory system and by irritability of the central nervous system. NAS often requires detoxification treatment in the hospital with a morphine taper. Reported rates of withdrawal symptoms among neonates born to opioid-addicted mothers who continued to use opiates within a week of giving birth range from 55% to 94%, and rates of NAS that develop among neonates as a result of treating the mother with MMT during pregnancy fall into this range. Recent studies on the long-term impact of NAS on development are scant. Older studies indicated no differences in cognitive performance among four-year old children of mothers receiving MMT and children of mothers with similar demographic characteristics in a control group. However, scores of children in both groups were lower than population norms." "Between 1999 and 2004, deaths attributed to methadone increased by 390%. Evidence suggests that this change was largely related to the increased use of methadone for pain analgesia rather than MMT. Nonetheless, the sharp rise of methadone-related deaths highlights safety issues—in particular, the risks of respiratory depression and cardiac QT interval prolongation. The QT interval is a measure of time between the start of the Q wave and the end of the T wave in the heart's electrical cycle that is measured by an electrocardiogram. Prolongation of the QT interval can lead to serious heart arrhythmias such as Torsades de Pointes (TdP) and sudden death."

Systematic Review	Evidence
	"As a result of this rise in mortality, the U.S. Food and Drug Administration issued a physician safety alert in 2006 highlighting fatalities and cardiac arrhythmias associated with methadone. Respiratory depression is most often a consequence of methadone accumulation and use of concurrent illicit drugs or medications that also suppress the central nervous system. Reviews suggest that initiation into methadone treatment is a particularly vulnerable time in both methadone maintenance and pain therapy populations, particularly if the dose is increased rapidly. The most common drugs associated with respiratory suppression are benzodiazepines and alcohol. Deaths from respiratory depression may also be caused by inappropriate dosing by methadone recipients and by diversion of methadone, which occurs when individuals who have a prescription for methadone sell or give their methadone to others rather than using it themselves."
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	See systematic review 1 which supports these recommendations.

*cell intentionally left blank

Systematic Review	Evidence
Source of Systematic Review 3: • Title • Author • Date • Citation, including page number • URL	Title: Medication-Assisted Treatment With Buprenorphine: Assessing the Evidence Author: Thomas, C. P., Fullerton, C. A., Kim, M., Montejano, L., Lyman, D. R., Dougherty, R. H., Daniels, A., S., Ghose, S. S., Delphin-Rittmon, M. E. Date: 2014 Citation: Thomas, C. P., Fullerton, C. A., Kim, M., Montejano, L., Lyman, D. R., Dougherty, R. H., ... & Delphin-Rittmon, M. E. (2014). Medication-assisted treatment with buprenorphine: assessing the evidence. *Psychiatric Services*, *65*(2), 158-170. URL: N/A
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.	"[Buprenorphine Maintenance Treatment] (BMT)" is associated with improved outcomes compared with placebo for individuals and pregnant women with opioid use disorders." "Policy makers have reason to promote access to BMT for patients in substance use treatment who may wish to choose BMT as a potentially safer alternative to [Methadone Maintenance Treatment (MMT)]. Administrators of substance use treatment programs, community health centers, and managed care organizations and other purchasers of health

Systematic Review	Evidence
	care services, such as Medicare, Medicaid, and commercial insurance carriers, should give careful consideration to BMT as a covered benefit."
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The research designs of the identified studies were examined. Three levels of evidence (high, moderate, and low) were used to indicate the overall research quality of the collection of studies. Ratings were based on predefined benchmarks that considered the number of studies and their methodological quality. If ratings were dissimilar (occurring for 13% of the studies rated), a consensus opinion was reached." "In general, high ratings indicate confidence in the reported outcomes and are based on three or more RCTs with adequate designs or two RCTs plus two quasi-experimental studies with adequate designs. Moderate ratings indicate that there is some adequate research to judge the service, although it is possible that future research could influence reported results. Moderate ratings are based on the following three options: two or more quasiexperimental studies with adequate design; one quasi-experimental study plus one RCT with adequate design; or at least two RCTs with some methodological weaknesses or at least three quasi-experimental studies with some methodological weaknesses. Low ratings indicate that research for this service is not adequate to draw evidence based conclusions. Low ratings indicate that studies have nonexperimental designs, there are no RCTs, or there is no more than one adequately designed quasi-experimental study." The grade assigned to the evidence was "high". The author's stated that "because of the large number of trials, the overall evidence for BMT was rated as high. Thus the level of research evidence is similar for BMT and MMT. In addition, multiple meta-analyses, reviews, and more than three independent RCTs have compared BMT with MMT on the primary outcomes stated above, and these results are also based on a high level of evidence in RCTs or reviews. Secondary outcomes, such as use of other illicit drugs, criminal behaviors, and other measures of addiction severity or psychosocial functioning varied among studies; as a result, the evidence for these secondary outcomes is not as strong."
Provide all other grades and definitions from the evidence grading system	*
Grade assigned to the **recommendation** with definition of the grade	"Evidence for the effectiveness of BMT: high. Evidence clearly shows that BMT has a positive impact compared with placebo on: retention in treatment, and illicit opioid use. Evidence is mixed for its impact on: nonopioid illicit drug use." "In general, high ratings indicate confidence in the reported outcomes and are based on three or more RCTs with adequate designs or two RCTs plus two quasi-experimental studies with adequate designs. Moderate

Systematic Review	Evidence
	ratings indicate that there is some adequate research to judge the service, although it is possible that future research could influence reported results. Moderate ratings are based on the following three options: two or more quasiexperimental studies with adequate design; one quasi-experimental study plus one RCT with adequate design; or at least two RCTs with some methodological weaknesses or at least three quasi-experimental studies with some methodological weaknesses. Low ratings indicate that research for this service is not adequate to draw evidence based conclusions. Low ratings indicate that studies have nonexperimental designs, there are no RCTs, or there is no more than one adequately designed quasi-experimental study."
Provide all other grades and definitions from the recommendation grading system	*
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	"The literature search revealed 16 RCTs, a randomized cross-over study, a study using a self-administered survey, and a retrospective descriptive study. RCTs used either buprenorphine alone or buprenorphine-naloxone. The search also found seven reviews or meta-analyses."
Estimates of benefit and consistency across studies	"Buprenorphine has a better safety profile than methadone, and the ability to prescribe buprenorphine in office facilities as opposed to only in opioid treatment programs improves access to care and earlier initiation of treatment. A key advantage of buprenorphine is its availability." "Both BMT and MMT improve pregnancy-related outcomes by reducing illicit drug use during pregnancy."
What harms were identified?	"The pharmacology of buprenorphine affords it a better safety profile than methadone, which is important considering that methadone is associated with one-third of opioid-related overdose deaths annually. Because it is a partial agonist at the mu opiate receptor, it has a ceiling effect that limits its potential to cause respiratory depression compared with methadone. However, this risk still exists, especially if buprenorphine is used in combination with other central nervous system depressants such as benzodiazepines or alcohol or is used in higher doses. In addition, unlike methadone, buprenorphine at standard doses does not affect cardiac electrophysiology by lengthening the cardiac QT interval—a mechanism that can lead to serious cardiac arrhythmias. Buprenorphine also has fewer drug interactions than methadone, especially with HIV medications. Taken together, the articles reviewed suggest that the efficacy of BMT is dose dependent, and dose is important to take into account when comparing medications."

Systematic Review	Evidence
	"Infants of mothers treated with buprenorphine during pregnancy may be born with NAS, although NAS appears to be less severe in infants of mothers treated with buprenorphine than of those treated with methadone." "Buprenorphine naloxone retains some potential for abuse intravenously, but the combination has less abuse potential as measured by self-administration than buprenorphine alone or heroin."
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	See systematic review 1 which supports these recommendations.

*cell intentionally left blank

Systematic Review	Evidence
Source of Systematic Review 4: • Title • Author • Date • Citation, including page number • URL	Title: Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. Author: Mattick RP, Breen C, Kimber J, and Davoli M. Date: 2014 Citation: Mattick RP, Breen C, Kimber J, and Davoli M. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. Cochrane Database of Systematic Reviews 2014, Issue 2. Art. No.: CD002207. DOI: 10.1002/14651858.CD002207.pub4. URL: http://onlinelibrary.wiley.com/doi/-10.1002/14651858.CD002207.pub4/epdf
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.	A meta-analysis of 31 trials found that the quality of evidence for buprenorphine maintenance varied from high to moderate (R.P. Mattick et al., 2014). The analysis examined randomized controlled trials of buprenorphine maintenance treatment versus placebo or methadone treatment for management of opioid use disorders. Strong evidence indicated that buprenorphine is superior to placebo medication in retention of participants in treatment at all dosing levels considered in the analyses. Specifically, buprenorphine retained participants better than a placebo at low doses (2 to 6 mg), at medium doses (7 to 15 mg), and at high doses (≥ 16 mg). The authors based their conclusion on placebo-controlled trials, concluding that buprenorphine is an effective medication for retaining individuals with an OUD in treatment at any dose above 2 mg and for suppressing illicit opioid use (at doses 16 mg or greater).

Systematic Review	Evidence
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The Cochrane review meta-analyses include grades. The meta-analysis of buprenorphine treatment (R.P. Mattick et al., 2014) assigned high to moderate grades to the studies under review.
Provide all other grades and definitions from the evidence grading system	High quality: Further research is very unlikely to change the authors' confidence in the estimated effect. Moderate quality: Further research is likely to have an important impact on confidence in the estimate of effect and may change the estimate.
Grade assigned to the **recommendation** with definition of the grade	The Cochrane review meta-analyses include grades. The meta-analysis of buprenorphine treatment (R.P. Mattick et al., 2014) assigned high to moderate grades to the studies under review.
Provide all other grades and definitions from the recommendation grading system	High quality: Further research is very unlikely to change the authors' confidence in the estimated effect. Moderate quality: Further research is likely to have an important impact on confidence in the estimate of effect and may change the estimate.
Body of evidence: Quantity – how many studies? Quality – what type of studies?	The evidence from the buprenorphine meta-analysis derives from a search of databases from 2003 to 2013. 31 randomized controlled trials of buprenorphine maintenance treatment versus placebo or methadone treatment for management of opioid use disorders were included.
Estimates of benefit and consistency across studies	Strong evidence indicated that buprenorphine is superior to placebo medication in retention of participants in treatment at all dosing levels considered in the analyses. Specifically, buprenorphine retained participants better than a placebo at low doses (2 to 6 mg), at medium doses (7 to 15 mg), and at high doses (≥ 16 mg). The authors based their conclusion on placebo-controlled trials, concluding that buprenorphine is an effective medication for retaining individuals with an OUD in treatment at any dose above 2 mg and for suppressing illicit opioid use (at doses 16 mg or greater).
What harms were identified?	No harms were identified.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	See systematic reviews 3, 2, and 1, which support these recommendations.

Systematic Review	Evidence
Source of Systematic Review 5: • Title • Author • Date • Citation, including page number • URL	Source Title: Methadone maintenance therapy versus no opioid replacement therapy for opioid dependence Author: Mattick RP, Breen C, Kimber J, and Davoli M. Date: 2009 Citation: Mattick RP, Breen C, Kimber J, and Davoli M. Methadone maintenance therapy versus no opioid replacement therapy for opioid dependence. Cochrane Database of Systematic Reviews 2009, Issue 3. Art. No.: CD002209. URL: http://onlinelibrary.wiley.com/doi/-10.1002/14651858.CD002209.pub2/epdf
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.	Another meta-analysis included 11 studies that met the criteria for inclusion in a Cochrane review of methadone treatment. All the studies were randomized clinical trials, two were double-blind. Methadone appeared statistically significantly more effective than nonpharmacological approaches in retaining patients in treatment and in reducing heroin use as measured by self-report and urine/hair analysis. The authors concluded that methadone is an effective maintenance therapy intervention for the treatment of OUD as it retains patients in treatment and decreases heroin use better than treatments that do not use opioid replacement therapy. However, the authors did not show a statistically significant superior effect on criminal activity or mortality.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	High to Moderate
Provide all other grades and definitions from the evidence grading system	High quality: Further research is very unlikely to change the authors' confidence in the estimated effect. Moderate quality: Further research is likely to have an important impact on confidence in the estimate of effect and may change the estimate.
Grade assigned to the **recommendation** with definition of the grade	*
Provide all other grades and definitions from the recommendation grading system	*
Body of evidence: Quantity – how many studies?	The evidence from the methadone meta-analysis derives from a search of databases from 2001 to 2008. 11 studies were included in the Cochrane Review. The quality of the evidence varied from high to moderate. The results from the 11 randomized trials all showed statistically significant

Systematic Review	Evidence
Quality – what type of studies?	positive benefits from methadone treatment, despite small sample sizes. All the studies were randomized clinical trials, and two were double-blind.
Estimates of benefit and consistency across studies	Methadone treatment appeared to be statistically significantly more effective than nonpharmacological approaches in retaining patients in treatment and reducing opioid use as measured by self-report and urine/hair analysis (six RCTs, RR 0.66, 95 percent CI 0.56 to 0.78), but not statistically different in reducing criminal activity (three RCTs, RR 0.39, 95 percent CI 0.12 to 1.25) or mortality (four RCTs, RR 0.48, 95 percent CI 0.10 to 2.39).
What harms were identified?	Given that none of the negative effects described above were statistically significant differences, this study did not identify any harms.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	See systematic reviews 4, 2, and 1, which support these recommendations.

*cell intentionally left blank

Systematic Review	Evidence
Source of Systematic Review 6: • Title • Author • Date • Citation, including page number • URL	Title: Medication-Assisted Treatment for Opioid Addiction in Opioid Treatment Programs. Treatment Improvement Protocol (TIP) Series 43 Author: Center for Substance Abuse Treatment Date: 2005 Citation: Center for Substance Abuse Treatment (2005). *Medication-Assisted Treatment for Opioid Addiction in Opioid Treatment Programs. Treatment Improvement Protocol (TIP) Series 43*. HHS Publication No. (SMA) 12-4214. Rockville, MD: Substance Abuse and Mental Health Services Administration. Page xvii. URL: https://www.ncbi.nlm.nih.gov/books-/NBK64164/pdf/Bookshelf_NBK64164.pdf
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize	p. xvii of executive summary, "Research supports the perspective that opioid addiction is a medical disorder that can be treated effectively with medications when they are administered under conditions consistent with their pharmacological efficacy and when treatment includes necessary supportive services such as psychosocial counseling, treatment for co-occurring disorders, medical services, and vocational rehabilitation."

Systematic Review	Evidence
the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Not Applicable. The guidelines cited above do not provide grades (e.g., USPSTF grades A, B, etc.).
Provide all other grades and definitions from the evidence grading system	A consensus panel of experts developed the SAMHSA Treatment Improvement Protocols (CSAT TIP 43). A team of external field reviewers then reviewed and commented on the guideline recommendations.
Grade assigned to the **recommendation** with definition of the grade	Not Applicable. The guidelines cited above do not provide grades (e.g., USPSTF grades A, B, etc.).
Provide all other grades and definitions from the recommendation grading system	A consensus panel of experts developed the SAMHSA Treatment Improvement Protocols (CSAT TIP 43). A team of external field reviewers then reviewed and commented on the guideline recommendations.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	The number of studies referenced in support of the guidelines was 608. The quality of the evidence selected indicates that included studies were considered to be of high quality by an expert panel.
Estimates of benefit and consistency across studies	Methadone maintenance is safe and effective, especially when used with psychosocial services (O'Connor & Fiellin, 2000). Maintenance treatment typically leads to reduction or cessation of illicit opioid use. A meta-analysis of 11 studies of the effectiveness of methadone (R.P. Mattick, Breen, Kimber, & Davoli, 2003) found that methadone was more effective than nonpharmacological treatment in retaining clients and reducing their opioid use. Clinical trials have demonstrated the primary efficacy of buprenorphine in patient retention as well as in the elimination of or reduction in opioid use (Fudala et al., 2003; Johnson, Strain, & Amass, 2003). Several studies evaluating the efficacy of buprenorphine for maintenance treatment lasting up to one year found that daily doses of 8 mg of sublingual solution or 8 to 16 mg of the buprenorphine tablet are safe and well tolerated. Most studies comparing buprenorphine and methadone have shown that 8 mg of sublingual buprenorphine or 16 mg of the tablet per day is equivalent to approximately 60 mg of oral methadone per day (Johnson et al., 2003). Naltrexone is effective in preventing relapse when used as directed; however, high rates of dropout have been reported. One study

Systematic Review	Evidence
	(Rothenberg et al., 2002) found especially poor retention among clients who had received methadone before naltrexone treatment. Naltrexone, under certain conditions, has resulted in better treatment compliance, e.g., when clients were supported with the opportunity to earn vouchers for treatment compliance (i.e., for each naltrexone dose ingested) (Preston et al., 1999). It should be noted, however, that CSAT TIP 43 predates approval of the long-acting naltrexone formulation for opioid use disorder and therefore does not address the effectiveness of the long-acting injectable preparation.
What harms were identified?	A small number of clients (10 percent) using naltrexone may experience gastrointestinal side effects that may necessitate their stopping the medication. Most clients, however, experience only mild, transient stomach upset. Some other side effects may include anxiety, nervousness, insomnia, headache, joint or muscle pain, and tiredness (Center for Substance Abuse Treatment, 2005).
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	See below systematic reviews 5, 4, 3, 2, and 1, which support these recommendations.

Systematic Review	Evidence
Source of Systematic Review 7 • Title • Author • Date • Citation, including page number • URL	Source Title: Clinical guidelines for the use of buprenorphine in the treatment of opioid addiction. Treatment Improvement Protocol (TIP) Series 40 Author: Center for Substance Abuse Treatment Date: 2004 Citation: Center for Substance Abuse Treatment (2004). *Clinical guidelines for the use of buprenorphine in the treatment of opioid addiction. Treatment Improvement Protocol (TIP) Series 40*. DHHS Publication No. (SMA) 04-3939. Rockville, MD: Substance Abuse and Mental Health Services Administration. Page 50. URL: https://www.ncbi.nlm.nih.gov/books-/NBK64245/pdf/BookshelfNBK6445.pdf
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome	p. 50, "The consensus panel recommends that the buprenorphine/naloxone combination be used for induction treatment (and for stabilization and maintenance) for most patients [with an OUD]."

Systematic Review	Evidence
being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Not Applicable. The guidelines cited above do not provide grades (e.g., USPSTF grades A, B, etc.).
Provide all other grades and definitions from the evidence grading system	A consensus panel of experts developed the SAMHSA Treatment Improvement Protocols (CSAT TIP 40). A team of external field reviewers then reviewed and commented on the guideline recommendations.
Grade assigned to the **recommendation** with definition of the grade	Not Applicable. The guidelines cited above do not provide grades (e.g., USPSTF grades A, B, etc.).
Provide all other grades and definitions from the recommendation grading system	A consensus panel of experts developed the SAMHSA Treatment Improvement Protocols (CSAT TIP 40). A team of external field reviewers then reviewed and commented on the guideline recommendations.
Body of evidence: Quantity – how many studies? Quality – what type of studies?	The number of studies referenced in support of the guidelines was 180. The quality of the evidence selected indicates that included studies were considered to be of high quality by an expert panel.
Estimates of benefit and consistency across studies	Clinical trials that compared buprenorphine to a placebo and to methadone have established the effectiveness of buprenorphine as a maintenance treatment of opioid addiction (Center for Substance Abuse Treatment, 2004). Buprenorphine treatment, compared to a placebo, is effective in reducing opioid use (Johnson et al., 1995). Evidence demonstrates that higher doses of buprenorphine and methadone are more effective in reducing opioid use (Ling et al., 1998; Petitjean et al., 2001; Schottenfeld, Pakes, Oliveto, Ziedonis, & Kosten, 1997). Another randomized trial found buprenorphine to be as effective as methadone, 60 mg/d, for retaining patients and reducing their opioid use. Both medications were superior to methadone at a lower level (20 mg/d) in reducing illicit opioid use and maintaining patients in treatment for 25 weeks (Johnson, Jaffe, & Fudala, 1992). A multisite office-based randomized study compared the effectiveness and safety of buprenorphine (16 mg) in combination with naloxone (4 mg), buprenorphine alone (16 mg), and a placebo (Fudala et al., 2003). The buprenorphine/naloxone in combination and buprenorphine alone demonstrated greater efficacy than the placebo. The proportion of urine

Systematic Review	Evidence
	samples that were negative for opiates was greater in the combined-treatment and buprenorphine-alone groups (17.8 and 20.7 percent, respectively) than in the placebo group (5.8 percent, $p < 0.001$ for both comparisons). Both buprenorphine treatment groups also reported significantly less opiate craving ($p < 0.001$ for both comparisons with placebo).
What harms were identified?	Buprenorphine and combinations of buprenorphine and naloxone are generally well tolerated, although side effects reported with these medications include headache, anxiety, constipation, perspiration, fluid retention in lower extremities, urinary hesitancy, and sleep disturbance (Center for Substance Abuse Treatment, 2004).
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	See below systematic reviews 6, 4, 3, and 1, which support these recommendations.

Systematic Review	Evidence
Source of Systematic Review 6: • Title • Author • Date • Citation, including page number • URL	Title: Medications for Opioid Use Disorder. Treatment Improvement Protocol (TIP) Series 63 Author: Substance Abuse and Mental Health Services Administration Date: 2020 Citation: Substance Abuse and Mental Health Services Administration. Medications for Opioid Use Disorder. Treatment Improvement Protocol (TIP) Series 63. Publication No. PEP20-02-01-006. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2020. URL: https://www.ncbi.nlm.nih.gov/books-/NBK64164/pdf/Bookshelf_NBK64164.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.	p. 1-5, "The TIP expert panel strongly recommends informing all patients with OUD about the risks and benefits of treatment of OUD with all FDA-Approved medications. Alternatives to these treatments and their risks and benefits should be discussed. Patients should receive access to such medications if clinically appropriate and desired by the patients." p. 1-8, "Patients can take medication for OUD on a short-term or long-term basis. However, patients who discontinue OUD medication generally return to illicit opioid use."

Systematic Review	Evidence
	p. 1-8, "The best results occur when a patient receives medication for as long as it provides a benefit. This approach is often called 'maintenance treatment.'"
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	Not Applicable. The guidelines cited above do not provide grades (e.g., USPSTF grades A, B, etc.).
Provide all other grades and definitions from the evidence grading system	A consensus panel of experts developed the SAMHSA Treatment Improvement Protocols (CSAT TIP 60). A team of external field reviewers then reviewed and commented on the guideline recommendations.
Grade assigned to the **recommendation** with definition of the grade	Not Applicable. The guidelines cited above do not provide grades (e.g., USPSTF grades A, B, etc.).
Provide all other grades and definitions from the recommendation grading system	A consensus panel of experts developed the SAMHSA Treatment Improvement Protocols (CSAT TIP 60). A team of external field reviewers then reviewed and commented on the guideline recommendations.
Body of evidence: Quantity – how many studies? Quality – what type of studies?	The number of studies referenced in support of the guidelines was 87. The quality of the evidence selected indicates that included studies were considered to be of high quality by an expert panel.
Estimates of benefit and consistency across studies	Multiple studies have demonstrated that methadone, extended-release injectable naltrexone (XR-NTX), and buprenorphine were each found to be more effective in reducing illicit opioid use than no medication in randomized clinical trials. Additional studies indicated that methadone and buprenorphine treatment have also been associated with reduced risk of overdose death.
What harms were identified?	Buprenorphine has a very low risk of medication-induced respiratory depression, possible risk of precipitated withdrawal when starting medication if started too prematurely after recent use of other opioids, and side effects of constipation, vomiting, headache, sweating, insomnia, and blurred vision. Methadone has higher risk of medication-induced respiratory depression than buprenorphine, although still rare. It has no risk of precipitated withdrawal when starting medication, and side effects include constipation, vomiting, sweating, dizziness, and sedation. Naltrexone has no risk of medication-induced respiratory depression, although severe withdrawal when starting medication is possible if period of abstinence is inadequate before starting medication. Common side effects include difficulty sleeping, anxiety, nausea, vomiting, low energy, joint and muscle pain, headache, and liver enzyme elevation.

Systematic Review	Evidence
	Extended-release injectable naltrexone also can have the additional side effects of injection site pain, nasopharyngitis, insomnia, and toothache.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	*

*cell intentionally left blank

1a.4 OTHER SOURCE OF EVIDENCE

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

N.A.

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

In 2018, opioids were involved in 46,802 overdose deaths in 2018 in the United States. Provisional numbers from the CDC show that in 2019 were 49,912 opioid overdose deaths, a 6.6% annual increase. Successful medication (MOUD) initiation and retention are the two clinical milestones likely to have the greatest impact on lowering risk of mortality from opioid use disorders (Degenhardt et al., 2010; Sordo et al., 2017; Volkow et al., 2019). Many randomized clinical trials and observational studies find that individuals with an opioid use disorder (OUD) who are treated with an FDA-approved opioid use disorder medication (i.e., methadone, buprenorphine, naltrexone) have better outcomes than individuals who do not receive OUD medications including larger reductions in mortality, opioid usage and relapses, rates of infectious disease, and emergency department or inpatient admissions, as well as improved functioning in major life domains (Clark et al., 2015; Clark et al., 2014; Council of Economic Advisers, 2019; D'Onofrio et al., 2015; Fullerton et al., 2014; Haley et al., 2019; Jarvis et al., 2018; Larochelle et al., 2018; Lo-Ciganic et al., 2016; Ma et al., 2019; Mark et al., 2020; Mattick et al., 2009; Mattick et al., 2014; Parran et al., 2010; Pierce et al., 2016; Schwarz et al., 2012; Sordo et al., 2017; Syed & Keating, 2013; Thomas et al., 2014; Williams et al., 2020; Woody et al., 2014). As described below, comprehensive reviews conclude that the evidence for the effectiveness of MOUD relative to non-medication based OUD treatment is of high to moderate quality (American Society of Addiction Medicine, 2020; Fullerton et al., 2014; Thomas et al., 2014, Mattick et al., 2014, Mattick et al., 2009, Center for Substance Abuse Treatment, 2004, SAMHSA, 2020).

NQF endorsed Use of Pharmacotherapy (NQF #3400) – a measure of percentage of Medicaid beneficiaries ages 18 to 64 with an OUD who filled a prescription for or were administered or ordered an FDA-approved medication for opioid use disorder (MOUD) during the measure year - at the health plan/Medicaid program level. The implementation of this measure is hypothesized to lead to more people receiving MOUD, improved health care outcomes, higher treatment costs, but lower overall costs because the comorbidities associated with OUD are reduced (See Figure 1. Logic model) (Ball & Ross, 1991; Busch et al., 2017; Clark et al., 2015; Florence et al., 2013; Mohlman et al., 2016; National Institute of Drug Abuse, 2019; Nielsen et al., 2016; Ronquest et al., 2018). The potential untended

consequences of the measure may include diversion and accidental overdoses such as by children. A recent review of 17 studies finds that most people use illicit buprenorphine to self-medicate (to manage opioid withdrawal symptoms or achieve or maintain abstinence from other opioids), while a smaller percentage use it to get high (Chilcoat et al., 2019). Between 2004 and 2011 there were 5,222 emergency department visits by children ages 1 to 5 involving accidental ingestion of buprenorphine (Crane, 2017). To put that number in perspective, there were 8.2 million prescriptions for buprenorphine in 2012. Buprenorphine is available in an extended release injectable form which has no diversion potential and no potential for accidental ingestion.

Despite the clear benefits of MOUD, many practitioners do not offer them to their patients with OUD. In 2018, only 40% of specialty addiction treatment facilities offered medications to treat opioid use disorder (Mark et al., 2020). In 2015, between 31% and 37% of patients with OUD in specialty facilities received medications for OUD (SAMHSA, 2017). In many regions of the county lack adequate supply of buprenorphine waived professionals (Abraham et al., 2020, Andrilla et al., 2020). A recent study in Massachusetts reported that only 30% of those who survived an opioid overdose received medications for OUD in the year after their overdose (Larochelle, 2018).

A 2014 American Society of Addiction Medicine expert panel recommended that the Use of Pharmacotherapy Measure (NQF #3400) be created as a measure at the clinician level (ASAM, 2014). They note that individual providers use measures for internal quality improvement and to monitor their practices (ASAM, 2014). The ASAM expert panel noted that the measure would be useful despite the fact that there is no set time in which a patient with OUD should be initiated on MOUD or an absolute benchmark level to determine the appropriate percentages of patients on MOUD. They explain that ensuring timely treatment with OUD medications is important as data show that individuals who receive MOUD are more likely to be retained in addiction treatment (Mattick et al., 2014, Timko et al., 2016). Further, improved outcomes have been demonstrated for patients who initiate buprenorphine to treat OUD in emergency departments (D'Onofrio et al., 2015). Finally, as noted above, a large and robust portfolio of research highlights the significant mortality and morbidity benefits of MOUD.

A measure of use of MOUD for OUD was developed and tested at the provider level in four states using Medicaid claims data and commercial claims data as part of a project sponsored by the nonprofit Shatterproof and conducted by RTI international. They study found that the measure was reliable and valid (Dowd et al., 2020).

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

PubMed searches were conducted using keywords: opioid use disorder, buprenorphine, methadone, naltrexone, vivitrol, and medication assisted therapy. The reference lists of articles were reviewed to identify articles that might have not been identified in the PubMed search.

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

1. Abraham AJ, Andrews CM, Harris SJ, Friedmann PD. Availability of Medications for the Treatment of Alcohol and Opioid Use Disorder in the USA. Neurotherapeutics. 2020;17(1):55-69. doi:10.1007/s13311-019-00814-4
2. American Society of Addiction Medicine. (2015). National practice guideline for the use of medications in the treatment of addiction involving opioid use. Retrieved from http://www.asam.org/quality-practice/guidelines-and-consensus-documents/npg/jam-article

3. American Society of Addiction Medicine (2020). National practice guideline for the treatment of opioid use disorder: 2020 focused update
4. American Society of Addiction Medicine (2014). The ASAM Performance Measures for the Addiction Specialist Physician. Retrieved from https://www.asam.org/docs/default-source/advocacy/performance-measures-for-the-addiction-specialist-physician.pdf
5. Andrilla CHA, Moore TE, Patterson DG, Larson EH. Geographic Distribution of Providers With a DEA Waiver to Prescribe Buprenorphine for the Treatment of Opioid Use Disorder: A 5-Year Update. J Rural Health. 2019;35(1):108-112. doi:10.1111/jrh.12307
6. Ball, J. C., & Ross, A. (1991). *The Effectiveness of Methadone Maintenance Treatment: Patients, Programs, Services, and Outcome*. New York: Springer Verlag.
7. Busch, S. H., Fiellin, D. A., Chawarski, M. C., Owens, P. H., Pantalon, M. V., Hawk, K., Bernstein, S. L., O'Connor, P. G., & D'Onofrio, G. (2017). Cost-effectiveness of emergency department-initiated treatment for opioid dependence: Treatment for opioid dependence. Addiction, 112(11), 2002–2010. https://doi.org/10.1111/add.13900
8. Center for Substance Abuse Treatment. (2004). *Clinical guidelines for the use of buprenorphine in the treatment of opioid addiction. Treatment Improvement Protocol (TIP) Series 40. DHHS Publication No. (SMA) 04-3939.* . Retrieved from Rockville, MD: Substance Abuse and Mental Health Services Administration: https://www.ncbi.nlm.nih.gov/books/NBK64245/
9. Center for Substance Abuse Treatment. (2005). *Medication-Assisted Treatment for Opioid Addiction in Opioid Treatment Programs. Treatment Improvement Protocol (TIP) Series 43*. Retrieved from HHS Publication No. (SMA) 12-4214. Rockville, MD: Substance Abuse and Mental Health Services Administration.
10. CMS, CMCS Information Bulletin. New Reporting Measures for Substance Use Disorder (SUD)-focused Health Homes, November 17, 2019. https://www.medicaid.gov/federal-policy-guidance/downloads/cib112719.pdf
11. Chilcoat HD, Amick HR, Sherwood MR, Dunn KE. Buprenorphine in the United States: Motives for abuse, misuse, and diversion. J Subst Abuse Treat. 2019;104:148-157. doi:10.1016/j.jsat.2019.07.005
12. Clark, R. E., Baxter, J. D., Aweh, G., O'Connell, E., Fisher, W. H., & Barton, B. A. (2015). Risk Factors for Relapse and Higher Costs Among Medicaid Members with Opioid Dependence or Abuse: Opioid Agonists, Comorbidities, and Treatment History. *J Subst Abuse Treat, 57*, 75-80. doi:10.1016/j.jsat.2015.05.001
13. Clark, R. E., Baxter, J. D., Barton, B. A., Aweh, G., O'Connell, E., & Fisher, W. H. (2014). The impact of prior authorization on buprenorphine dose, relapse rates, and cost for Massachusetts Medicaid beneficiaries with opioid dependence. *Health Serv Res, 49*(6), 1964-1979. doi:10.1111/1475-6773.12201
14. Council of Economic Advisers. (2019). *The Full Cost of the Opioid Crisis: $2.5 Trillion Over Four Years*. https://www.whitehouse.gov/articles/full-cost-opioid-crisis-2-5-trillion-four-years/
15. Crane, E.H. Emergency Department Visits Involving the Accidental Ingestion of Opioid Pain Relievers by Children Aged 1 to 5. The CBHSQ Report: November 30, 2017. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, Rockville, MD.
16. Degenhardt L, Bucello C, Mathers B, Briegleb C, Ali H, Hickman M, McLaren J. Mortality among regular or dependent users of heroin and other opioids: a systematic review and meta-analysis of cohort studies. Addiction. 2011 Jan;106(1):32-51. doi: 10.1111/j.1360-0443.2010.03140.x. Epub 2010 Nov 4.
17. D'Onofrio, G., O'Connor, P. G., Pantalon, M. V., Chawarski, M. C., Busch, S. H., Owens, P. H., Bernstein, S. L., & Fiellin, D. A. (2015). Emergency Department–Initiated Buprenorphine/Naloxone Treatment for Opioid Dependence: A Randomized Clinical Trial. *JAMA*, 313(16), 1636. https://doi.org/10.1001/jama.2015.3474

18. Dowd, W.N., Mark, T.L., Barch, D.H., Popovic, J.R., Seibert, J.H. (2020). The Reliability and Validity of Provider-Level Claims-based Substance Use Disorder Quality Measures. Research Triangle Park, NC: RTI International
19. Florence, C. S., Zhou, C., Luo, F., & Xu, L. (2016). The Economic Burden of Prescription Opioid Overdose, Abuse, and Dependence in the United States, 2013: *Medical Care*, 54(10), 901–906. https://doi.org/10.1097/MLR.0000000000000625
20. Fudala, P. J., Bridge, T. P., Herbert, S., Williford, W. O., Chiang, C. N., Jones, K., . . . Buprenorphine/Naloxone Collaborative Study, G. (2003). Office-based treatment of opiate addiction with a sublingual-tablet formulation of buprenorphine and naloxone. *N Engl J Med, 349*(10), 949-958. doi:10.1056/NEJMoa022164
21. Fullerton, C. A., Kim, M., Thomas, C. P., Lyman, D. R., Montejano, L. B., Dougherty, R. H., . . . Delphin-Rittmon, M. E. (2014). Medication-assisted treatment with methadone: assessing the evidence. *Psychiatr Serv, 65*(2), 146-157. doi:10.1176/appi.ps.201300235
22. Haley, S. J., Maroko, A. R., Wyka, K., & Baker, M. R. (2019). The association between county-level safety net treatment access and opioid hospitalizations and mortality in New York. *Journal of Substance Abuse Treatment*, 100, 52–58. https://doi.org/10.1016/j.jsat.2019.02.004
23. Jarvis, B. P., Holtyn, A. F., Subramaniam, S., Tompkins, D. A., Oga, E. A., Bigelow, G. E., & Silverman, K. (2018). Extended-release injectable naltrexone for opioid use disorder: A systematic review: Review of XR-NTX for OUD. *Addiction*, 113(7), 1188–1209. https://doi.org/10.1111/add.14180
24. Johnson, R. E., Eissenberg, T., Stitzer, M. L., Strain, E. C., Liebson, I. A., & Bigelow, G. E. (1995). A placebo controlled clinical trial of buprenorphine as a treatment for opioid dependence. *Drug Alcohol Depend, 40*(1), 17-25.
25. Johnson, R. E., Jaffe, J. H., & Fudala, P. J. (1992). A controlled trial of buprenorphine treatment for opioid dependence. *JAMA, 267*(20), 2750-2755.
26. Johnson, R. E., Strain, E. C., & Amass, L. (2003). Buprenorphine: how to use it right. *Drug Alcohol Depend, 70*(2 Suppl), S59-77.
27. Larochelle, M. R., Bernson, D., Land, T., Stopka, T. J., Wang, N., Xuan, Z., Bagley, S. M., Liebschutz, J. M., & Walley, A. Y. (2018). Medication for Opioid Use Disorder After Nonfatal Opioid Overdose and Association With Mortality: A Cohort Study. *Annals of Internal Medicine*, 169(3), 137. https://doi.org/10.7326/M17-3107
28. Lo-Ciganic WH, Gellad WF, Gordon AJ, et al. Association between trajectories of buprenorphine treatment and emergency department and in-patient utilization. *Addiction*. 2016;111(5):892-902. doi:10.1111/add.13270
29. Ling, W., Charuvastra, C., Collins, J. F., Batki, S., Brown, L. S., Jr., Kintaudi, P., . . . Segal, D. (1998). Buprenorphine maintenance treatment of opiate dependence: a multicenter, randomized clinical trial. *Addiction, 93*(4), 475-486.
30. Ma, J., Bao, Y.-P., Wang, R.-J., Su, M.-F., Liu, M.-X., Li, J.-Q., Degenhardt, L., Farrell, M., Blow, F. C., Ilgen, M., Shi, J., & Lu, L. (2019). Effects of medication-assisted treatment on mortality among opioids users: A systematic review and meta-analysis. *Molecular Psychiatry*, 24(12), 1868–1883. https://doi.org/10.1038/s41380-018-0094-5
31. Mark, T. L., Parish, W. J., & Zarkin, G. A. (2020). Association of Formulary Prior Authorization Policies With Buprenorphine-Naloxone Prescriptions and Hospital and Emergency Department Use Among Medicare Beneficiaries. *JAMA Network Open*, 3(4), e203132. https://doi.org/10.1001/jamanetworkopen.2020.3132
32. Mark, T.L, Dowd, W.N , Council, C..L. Tracking the Qualty of Addiction Treatment Over Time and Across States: Using the Federal Government's "Signs" of Higher Quality. RTI Press. June, 2020

33. Mattick, R. P., Breen, C., Kimber, J., & Davoli, M. (2003). Methadone maintenance therapy versus no opioid replacement therapy for opioid dependence (Cochrane Review). *Cochrane Database Systems Review 2003(2):CD00209.*

34. Mattick, R. P., Breen, C., Kimber, J., & Davoli, M. (2009). *Methadone maintenance therapy versus no opioid replacement therapy for opioid dependence. Cochrane Database of Systematic Reviews 2009, Issue 3. Art. No.: CD002209. DOI: 10.1002/14651858.CD002209.pub2.* Retrieved from

35. Mattick, R. P., Breen, C., Kimber, J., & Davoli, M. (2014). Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD002207.pub4

36. Minozzi, S., Amato, L., Vecchi, S., Davoli, M., Kirchmayer, U., & Verster, A. (2011). Oral naltrexone maintenance treatment for opioid dependence. The Cochrane database of systematic reviews, 2011(4), CD001333. https://doi.org/10.1002/14651858.CD001333.pub4

37. Mohlman, M. K., Tanzman, B., Finison, K., Pinette, M., & Jones, C. (2016). Impact of Medication-Assisted Treatment for Opioid Addiction on Medicaid Expenditures and Health Services Utilization Rates in Vermont. *J Subst Abuse Treat, 67*, 9-14. doi:10.1016/j.jsat.2016.05.002

38. National Institute of Drug Abuse (NIDA). (2018, June 8). Medications to Treat Opioid Use Disorder. Retrieved from https://www.drugabuse.gov/publications/research-reports/medications-to-treat-opioid-use-disorder on 2020, May 11.

39. Nielsen, S., Larance, B., Degenhardt, L., Gowing, L., Kehler, C., & Lintzeris, N. (2016). Opioid agonist treatment for pharmaceutical opioid dependent people. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD011117.pub2

40. O'Connor, P. G., & Fiellin, D. A. (2000). Pharmacologic treatment of heroin-dependent patients. *Ann Intern Med, 133*(1), 40-54.

41. Parran, T. V., Adelman, C. A., Merkin, B., Pagano, M. E., Defranco, R., Ionescu, R. A., & Mace, A. G. (2010). Long-term outcomes of office-based buprenorphine/naloxone maintenance therapy. *Drug Alcohol Depend, 106*(1), 56-60. doi:10.1016/j.drugalcdep.2009.07.013

42. Petitjean, S., Stohler, R., Deglon, J. J., Livoti, S., Waldvogel, D., Uehlinger, C., & Ladewig, D. (2001). Double-blind randomized trial of buprenorphine and methadone in opiate dependence. *Drug Alcohol Depend, 62*(1), 97-104.

43. Pierce, M., Bird, S. M., Hickman, M., Marsden, J., Dunn, G., Jones, A., & Millar, T. (2016). Impact of treatment for opioid dependence on fatal drug-related poisoning: a national cohort study in England. *Addiction, 111*(2), 298-308. doi:10.1111/add.13193

44. Preston, K. L., Silverman, K., Umbricht, A., DeJesus, A., Montoya, I. D., & Schuster, C. R. (1999). Improvement in naltrexone treatment compliance with contingency management. *Drug Alcohol Depend, 54*(2), 127-135.

45. Ronquest, N., Willson, T., Montejano, L., Nadipelli, V., & Wollschlaeger, B. (2018). Relationship between buprenorphine adherence and relapse, health care utilization and costs in privately and publicly insured patients with opioid use disorder. *Substance Abuse and Rehabilitation, Volume 9*, 59–78. https://doi.org/10.2147/SAR.S150253

46. Rothenberg, J. L., Sullivan, M. A., Church, S. H., Seracini, A., Collins, E., Kleber, H. D., & Nunes, E. V. (2002). Behavioral naltrexone therapy: an integrated treatment for opiate dependence. *J Subst Abuse Treat, 23*(4), 351-360.

47. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. Treatment Episode Data Set (TEDS): 2005-2015. National Admissions to Substance Abuse Treatment Services. BHSIS Series S-91, HHS Publication No. (SMA) 17-5037. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2017.

48. Schottenfeld, R. S., Pakes, J. R., Oliveto, A., Ziedonis, D., & Kosten, T. R. (1997). Buprenorphine vs methadone maintenance treatment for concurrent opioid dependence and cocaine abuse. *Arch Gen Psychiatry, 54*(8), 713-720.

49. Schwarz, R., Zelenev, A., Bruce, R. D., & Altice, F. L. (2012). Retention on buprenorphine treatment reduces emergency department utilization, but not hospitalization, among treatment-seeking patients with opioid dependence. *Journal of Substance Abuse Treatment*, 43(4), 451–457. https://doi.org/10.1016/j.jsat.2012.03.008

50. Sordo, L., Barrio, G., Bravo, M. J., Indave, B. I., Degenhardt, L., Wiessing, L., Ferri, M., & Pastor-Barriuso, R. (2017). Mortality risk during and after opioid substitution treatment: Systematic review and meta-analysis of cohort studies. BMJ, j1550. https://doi.org/10.1136/bmj.j1550

51. Substance Abuse and Mental Health Services Administration. (2015). *Drug Addiction Treatment Act of 2000 (DATA 2000).* Retrieved from https://www.samhsa.gov/medication-assisted-treatment/legislation-regulations-guidelines#DATA-2000 Substance Abuse and Mental Health Services Administration. *Medications for Opioid Use Disorder*. Treatment Improvement Protocol (TIP) Series 63. Publication No. PEP20-02-01-006. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2020.

52. Syed, Y. Y., & Keating, G. M. (2013). Extended-Release Intramuscular Naltrexone (VIVITROL®): A Review of Its Use in the Prevention of Relapse to Opioid Dependence in Detoxified Patients. *CNS Drugs*, 27(10), 851–861. https://doi.org/10.1007/s40263-013-0110-x

53. Thomas, C. P., Fullerton, C. A., Kim, M., Montejano, L., Lyman, D. R., Dougherty, R. H., . . . Delphin-Rittmon, M. E. (2014). Medication-assisted treatment with buprenorphine: assessing the evidence. *Psychiatr Serv, 65*(2), 158-170. doi:10.1176/appi.ps.201300256

54. Timko C, Schultz NR, Cucciare MA, Vittorio L, Garrison-Diehn C. Retention in medication-assisted treatment for opiate dependence: A systematic review. J Addict Dis. 2016;35(1):22-35.

55. Volkow ND, Jones EB, Einstein EB, Wargo EM. Prevention and Treatment of Opioid Misuse and Addiction: A Review. JAMA Psychiatry. 2019;76(2):208-216. doi:10.1001/jamapsychiatry.2018.3126

56. Williams, A. R., Samples, H., Crystal, S., & Olfson, M. (2020). Acute Care, Prescription Opioid Use, and Overdose Following Discontinuation of Long-Term Buprenorphine Treatment for Opioid Use Disorder. *American Journal of Psychiatry*, 177(2), 117–124. https://doi.org/10.1176/appi.ajp.2019.19060612

57. Woody, G. E., Bruce, D., Korthuis, P. T., Chhatre, S., Poole, S., Hillhouse, M., Jacobs, P., Sorensen, J., Saxon, A. J., Metzger, D., & Ling, W. (2014). HIV risk reduction with buprenorphine-naloxone or methadone: findings from a randomized trial. *Journal of Acquired Immune Deficiency Syndromes* (1999), 66(3), 288–293. https://doi.org/10.1097/QAI.0000000000000165

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.

In 2019, opioids were involved in 49,912 overdose deaths in the United States, a 6.6% annual increase from 2018. The two clinical milestones likely to have the greatest impact on lowering the risk of mortality from opioid use disorders are initiation of medications to treat opioid use disorder (MOUD) initiation and retention on MOUD (Ball & Ross, 1991; Degenhardt et al., 2011; Sordo et al., 2017; Volkow et al., 2019). Randomized

clinical trials and observational studies find that individuals with an opioid use disorder (OUD) who are treated with an FDA-approved opioid use disorder medication (i.e., methadone, buprenorphine, naltrexone) have better outcomes than individuals who do not receive OUD medications such as larger reductions in mortality, opioid usage and relapses, rates of infectious disease, and emergency department and inpatient admissions, as well as improved functioning in major life domains (Clark et al., 2014; Clark et al., 2015; D´Onofrio et al., 2015; Fullerton et al., 2014; Haley et al., 2019; Jarvis et al., 2018; Larochelle et al., 2018; Lo-Ciganic et al., 2016; Ma et al., 2019; Mark et al., 2020; Mattick et al., 2003; Mattick et al., 2009; Mattick et al., 2014; Minozzi et al., 2011; Parran et al., 2010; Pierce et al., 2016; Schwarz et al., 2012; Sordo et al., 2017; Syed & Keating, 2013; Thomas et al., 2014; Williams et al., 2020; Woody et al., 2014). As described below, comprehensive reviews conclude that the evidence for the effectiveness of MOUD relative to non-medication based OUD treatment is of high to moderate quality (American Society of Addiction Medicine, 2015, 2020; Center for Substance Abuse Treatment, 2004, Center for Substance Abuse Treatment, 2005; Fullerton et al., 2014; Thomas et al., 2014, Mattick et al., 2014, Mattick et al., 2009, SAMHSA, 2020).

NQF endorsed Use of Pharmacotherapy (NQF #3400) – a measure of the percentage of Medicaid beneficiaries ages 18 to 64 with an OUD who filled a prescription for or were administered or ordered an FDA-approved medication for opioid use disorder (MOUD) during the measure year - at the health plan and Medicaid program level. The implementation of this measure is hypothesized to lead to more people receiving MOUD, improved health care outcomes, higher treatment costs, but lower overall costs because high-cost acute care utilization is reduced and the comorbidities associated with OUD are treated more effectively (See Figure 1. Logic model) (Busch et al., 2017; Clark et al., 2014; Clark et al., 2015; Florence et al., 2013; Mohlman et al., 2016; Mark et al., 2020; National Institute of Drug Abuse, 2019; Nielsen et al., 2016; Ronquest et al., 2018). The potential untended consequences of the measure may include diversion and accidental overdoses, such as by children. A recent review of 17 studies finds that most people use illicit buprenorphine to self-medicate (to manage opioid withdrawal symptoms or achieve or maintain abstinence from other opioids), while a smaller percentage use it to get high (Chilcoat et al., 2019). Between 2004 and 2011, there were 5,222 emergency department visits by children ages 1 to 5 involving accidental ingestion of buprenorphine (Crane, 2017). To put that number in perspective, there were 8.2 million prescriptions for buprenorphine in 2012. Buprenorphine is available in an extended-release injectable form with no diversion potential and no potential for accidental ingestion.

Despite the clear benefits of MOUD, many practitioners do not offer them to their patients with OUD. In 2018, only 40% of specialty addiction treatment facilities offered medications to treat opioid use disorder (Mark et al., 2020). In 2015, between 31% and 37% of patients with OUD in specialty facilities received medications for OUD (SAMHSA, 2017). Many regions of the country lack an adequate supply of buprenorphine waived professionals (Abraham et al., 2020; Andrilla et al., 2020). A recent study in Massachusetts reported that only 30% of those who survived an opioid overdose received medications for OUD in the year after their overdose (Larochelle, 2018).

A 2014 American Society of Addiction Medicine expert panel recommended that the Use of Pharmacotherapy Measure (NQF #3400) be created as a measure at the clinician level (ASAM, 2014). They note that individual providers could use the measure for quality improvement and to monitor their practices (ASAM, 2014). The ASAM expert panel noted that the measure would be useful even though there is no set time when a patient with OUD should be initiated on MOUD or an absolute benchmark level to determine the appropriate percentages of patients on MOUD. They explain that ensuring timely treatment with OUD medications is important as data show that individuals who receive MOUD are more likely to be retained in addiction treatment (Mattick et al., 2014, Timko et al., 2016). Further, improved outcomes have been demonstrated for patients who initiate buprenorphine to treat OUD in emergency departments (D'Onofrio et al., 2015). Finally, as noted above, a large and robust portfolio of research highlights the significant mortality and morbidity benefits of MOUD.

A measure of use of MOUD for OUD is being reported and used at the provider level in four state Medicaid programs to help identify providers who could benefit from technical assistance (New York, Massachusetts,

West Virginia, Delaware). The measure is also being used at the provider-level in Centers for Medicare and Medicaid (CMS) behavioral health home demonstrations; however, it has not been endorsed at the provider level (CMS, 2019).

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.*

We tested the measure using 2014 Medicaid Analytic Extract (MAX) data on 9398 SUD service providers (including both individual providers and facilities/group practices) that treated at least ten patients with an OUD diagnosis. The number of beneficiaries only eligible for Medicaid (not both Medicaid and Medicare) with at least one medical claim for an encounter with a primary or secondary OUD diagnosis in the calendar year was 716,431. The mean provider-level score was 38.4% with provider-level scores ranging from 0% to 100%. Below we present the provider-level score distribution.

Summary Data of Observed Scores

n	Mean	SD	Min	10th	25th	50th	75th	90th	Max
All Provider Types									
9398	38%	33%	0%	0%	10%	28%	66%	92%	100%
Individual Clinicians									
5344	44%	35%	0%	0%	10%	37%	79%	93%	100%
Hospitals/Facilities/Agencies									
4054	31%	28%	0%	3%	9%	22%	45%	80%	100%

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

Not applicable

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

There are disparities in the use of medications to treat opioid use disorder as revealed in the table below. The table describes the percentage of patients diagnosed with an opioid use disorder who received medications to treat opioid use disorder by gender and race. Men are less likely to receive medications than women. Blacks are less likely to receive medications than Whites. Hispanics/Latinos are more likely to receive medications than Whites or Blacks.

	Mean	N
Gender		
Male	43%	368,554
Female	48%	347,877
Total	46%	716,431
Race		
White	49%	463,846
Black	31%	68,414
American Indian/Alaskan Native	50%	7,290
Asian	42%	2,707
Hispanic/Latino	64%	23,957
Native Hawaiian/Pacific Islander	64%	1,881
Total	46%	716,431

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

Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

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

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

Not applicable

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

This is not an eMeasure **Attachment:**

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

Attachment**:** Data_Dictionary_for_MAT_Receipt_Measure__7-9-20.xlsx

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

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

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

Not an instrument-based measure

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

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

Not applicable.

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

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

Beneficiaries ages 18 to 64 with an OUD who filled a prescription for, or were administered or ordered, an FDA-approved medication for the treatment of OUD within 30 days of the first attributable encounter with an OUD diagnosis with the provider.

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).

The measure numerator is the number of beneficiaries ages 18 to 64 with an OUD diagnosis (see Appendix A in Data Dictionary) who filled a prescription for, or were administered or ordered, an FDA-approved medication to treat OUD (see Appendix B in Data Dictionary) within 30 days of the first attributable encounter with the provider.

Note that the OUD medication administration or prescription can be from any provider (e.g., office-based physician, hospital, OTP), it need not necessarily be the attributed provider. This justification is that all providers who treat patients with an OUD diagnosis should be held accountable for ensuring that they receive gold standard treatment.

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

Number of Medicaid ages 18 – 64 beneficiaries with at least one medical claim for an encounter with an OUD diagnosis with that provider (where the provider is identified by a National Provider Identifier (NPI) code).

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

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

The target population for the denominator includes all Medicaid beneficiaries aged 18 through 64 years with a diagnosis of OUD (primary or other) that had an encounter with the provider at least once during the measure

time period which is defined as a calendar year. See Appendix A for ICD codes for identifying OUD. Age is calculated as of December 31st of the measurement year. Denominator exclusions are described below in 5.8.

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

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

Individuals under 18 are excluded. Rationale: There is limited evidence regarding the efficacy of MOUD for this population.

Individuals over 64 are excluded: Rationale: Most individuals over age 64 are covered under Medicare. Services covered by Medicare would not be capture in the Medicaid claims data and therefore follow-up treatment would be missed.

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

Instructions for the analytic file build, including denominator exclusion detail are included below.

Measurement Year: Calendar year 2014

Ages: 18 years and older as of December 31 of the measurement year. 64 or younger as of December 31 of the measurement year

Required Benefits: Medical, Chemical Dependency, and Pharmacy

Analytic File Inclusion Criteria Follow steps below.

1. Subset file to patients who had an OUD diagnosis. (Appendix A contains ICD codes for identifying OUD) in any diagnostic position from any provider during selected Calendar year.
2. Eliminate dual eligible (Medicare/Medicaid) beneficiaries.
3. Eliminate any patient IDs of patients younger than 18 as of December 31 of the measurement year, or older than 64 as of December 31 of the measurement year.
4. Pull all the claims/records from the enrollment, inpatient, outpatient, prescription drug files, and long-term claims files with these Member IDs into an analytic sample.

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

Not applicable.

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = 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.)*

Identify denominator

Identify Medicaid beneficiaries aged 18 through 64 years with at least one encounter with a provider with an OUD diagnosis on the claim (primary or other secondary) during the measurement year. Must be continuously enrolled for at least 30 days after the attributable encounter . Age is calculated as of December 31st of the measurement year.

Step 1. Identify the attribution date, the first encounter between a member and a provider. The attribution date is as follows.

a. Outpatient Encounter Attribution Date: Attribution date is the date of the encounter with an outpatient provider that includes an OUD diagnosis (primary or secondary).
b. Inpatient/Residential Encounter Attribution Date. Attribution date is the discharge date from an inpatient/residential provider that includes an OUD diagnosis (any position).

Note: a member can be attributed to more than one provider at different times during the measurement period. However, if members have multiple attribution dates with a single provider, only the first is included in the denominator.

Step 2. Exclude a member from the denominator for a provider organization if the attribution date is after December 1 to allow for 30 days of time after the encounter.

Step 3. Only include members with continuous enrollment over the relevant 30 day time period.

Step 4. Exclude providers if their total number of attributable members is < 10.

Step 5. Count the number of patients in the denominator with a qualifying_medication_event_date (Appendix B) <= 30 days of the attribution_date (attribution_date <= qualifying_medication_event_date <= attribution_date + 30 days)

Step 6. Report measure metrics at the NPI level separately for individual clinicians and hospitals/agencies/facilities.

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

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

Not applicable. The measure is not based on a sample.

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

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

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

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

If other, please describe in S.18.

Claims, Enrollment Data

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

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

The data source is Medicaid Analytic Extract (MAX) files, including person summary (PS), inpatient (IP), other services (OT), long-term care (LT) and drug (RX) files. The other services file contains facility and individual provider services data. The Medicaid Analytic Extract (MAX) files contain data from 32 states.

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)*

Clinician : Individual, Facility

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.)*

Not applicable.

2. Validity – See attached Measure Testing Submission Form

MAT_Receipt_Attribution_Appendix_7_13_20.docx,MOUD_Receipt_Testing_Form_092720.docx

2.1 For maintenance of endorsement

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

2.2 For maintenance of endorsement

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

2.3 For maintenance of endorsement

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

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

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

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

Measure Title: Prescription or administration of pharmacotherapy for opioid use disorder (OUD)

Date of Submission:

Type of Measure:

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

*cell intentionally left blank

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

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

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

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

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

The Medicaid Analytic Extract (MAX) files were used to identify adult Medicaid beneficiaries ages 18 – 64 with at least one medical claim for an encounter with a primary or secondary opioid use disorder (OUD) diagnosis (denominator) and whether the patient received an OUD medication within 30 days of that encounter (numerator). Data came from 32 states.

The Medicaid MAX files used include the following types of files:

- **Person summary (PS).** Person-level file for Medicaid eligibility and demographic information.
- **Inpatient (IP).** Claim-level file for inpatient hospital stays.
- **Long-term care (LT**). Claim-level file for long-term care institutional stays (e.g., nursing facilities, intermediate care facilities for individuals with intellectual disabilities, psychiatric hospitals).
- **Other therapy (OT)**. Claim-level file for a wide variety of services, many of which are provided on an outpatient basis. Most notably, it may contain both residential and other stayover service claims data as claims are assigned to MAX claims file types based upon the category of service provided.
- **Rx file**. Claim-level file provides information on drugs and other services provided by a pharmacy.

Data from the PS IP, LT and OT files were used to construct the measure denominator. We used the PS file to limit the analytic sample based on age and enrollment criteria, and then we used the IP, LT, and OT files to determine whether those beneficiaries met the criteria for the measure's denominator. The IP, OT and Rx files enabled us to identify the numerator events (i.e., beneficiary filled a prescription for or were administered an FDA-approved medication for OUD within 30 days of the first attributable encounter with the provider). The PS file contained additional demographic and enrollment information, such as beneficiaries' state, disability status, age, sex, and race or ethnicity.

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

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

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

This measure is intended to be reported among all adult Medicaid beneficiaries at the provider level. A measure of prescription or administration of pharmacotherapy for OUD among individuals with an OUD diagnosis (NQF#3400) has been developed and endorsed at the population level (Population: State, Medicaid Program).

We identify a provider based on the National Provider Identifier (NPI) listed on the medical claims. A provider can be an individual clinician or a facility/group practice but excludes labs. A patient is attributed to a provider (NPI) if the provider submitted a claim for the patient for an encounter with an OUD diagnosis listed (primary or secondary position). Providers had to have at least 10 encounters with an OUD diagnosis to be included in the measure. Note that the OUD medication administration or prescription can be from any provider), it need not necessarily be the attributed provider. Further details of the attribution analysis are described in a supplemental document.

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	**Measure Tested at Level of:**
x☐ individual clinician	x☐ individual clinician
☐ group/practice	☐ group/practice

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

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

A total of 9,398 SUD providers comprising 5344 individual clinicians 4054 and hospitals/facilities/agencies.

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*)

We identified 716,431 beneficiaries only eligible for Medicaid (and not both Medicaid and Medicare) with at least one medical claim for an encounter with a primary or secondary OUD diagnosis (see **Table 2**). ICD codes were used to identify an OUD diagnosis. The codes used are listed in **Table 3** below. Slightly over half of the beneficiaries were male (51.4%) and about two-thirds were between the ages of 25 and 44 (62.4%). About two-thirds of the beneficiaries identified as White (64.7%) followed by Black (9.5%). Race or ethnicity was unknown for about eleven percent of beneficiaries (11.4%).

Table 2. Beneficiary Characteristics

Measures	Number of beneficiaries with OUD diagnosis (n)	Distribution of beneficiaries with OUD diagnosis (%)
Total	716,431	100.0
Medicaid-Category: Other[1]	31,546	4.4
Medicaid-Category: Aged	24	0.0

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

Measures	Number of beneficiaries with OUD diagnosis (n)	Distribution of beneficiaries with OUD diagnosis (%)
Medicaid-Category: Blind/disabled	158,598	22.1
Medicaid-Category: Adult	516,006	72.0
Medicaid-Category: Child	10,257	1.4
Age: 18-24	74,596	10.4
Age: 25-44	446,951	62.4
Age: 45-64	194,884	27.2
Gender: Male	368,554	51.4
Gender: Female	347,877	48.6
Race/ethnicity: White	463,846	64.7
Race/ethnicity: Black	68,414	9.5
Race/ethnicity: American Indian/Alaskan Native	7,290	1.0
Race/ethnicity: Asian	2,707	0.4
Race/ethnicity: Hispanic/Latino	23,957	3.3
Race/ethnicity: Native Hawaiian/Pacific Islander	1,881	0.3
Race/ethnicity: Other	66,694	9.3
Race/ethnicity: Unknown	81,642	11.4

Table 3. OUD Diagnosis Codes

ICD-10	Description
F11.1X	Opioid abuse
F11.2X	Opioid dependence1
ICD-9	*
304.0	Opioid type dependence
304.00	Opioid type dependence, unspecified
304.01	Opioid type dependence, continuous
304.02	Opioid type dependence, episodic
305.5	Opioid abuse
305.50	Opioid abuse
305.51	Opioid abuse, continuous
305.52	Opioid abuse, episodic
304.03	Opioid dependence in remission
304.53	Non-dependent opioid abuse in remission
304.7	Combinations of opioid type drug with any other drug dependence

ICD-10	Description
304.70	Combinations of opioid type drug with any other drug dependence, unspecified
304.71	Combinations of opioid type drug with any other drug dependence, continuous
304.72	Combinations of opioid type drug with any other drug dependence, episodic

[1] Including F11.21 (Opioid dependence, in remission).

*cell intentionally left blank

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

The same population was used for all aspects of testing.

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

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

2a2. RELIABILITY TESTING

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

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

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

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

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

Performance Score Reliability Testing

Reliability testing was conducted to determine whether the measure can distinguish well-performing providers from poorly performing providers. We conducted reliability tests for all provider types combined and independently for individual clinicians and hospitals/facilities/agencies.

To test reliability, we conducted three types of reliability tests:

1. **Signal-to-noise ratio and effect size tests**: We conducted parametric analysis of variance (ANOVA) and calculated effect size statistics. Both the observed F-ratios and the effect size statistics can be considered measures of signal-to-noise ratios where the signal is the effect created by true differences in underlying provider characteristics and the noise is all variance that is not explained by those differences. Effect size statistics describe the extent to which the independent variable (the provider) influenced the dependent variable (a "success" or "failure" for a patient on a particular measure). For a measure to be reliable, providers should have a demonstrable impact on patient outcomes as measured. For each analysis of variance, we calculate the *F* statistic and estimate two effect size statistics: (1) eta squared ($\eta 2$), which represents the proportion of the variance in the outcome that is explained by the provider, and (2) omega squared ($\omega 2$), which is similar to $\eta 2$ but is more robust with regard to small sample size. These tests provide insight into whether the provider has a significant effect on measure performance, as well as the size of that effect. An $\eta 2$ or $\omega 2$ squared value of 0.14 or higher indicates a large effect.
2. **Intra-unit reliability (IUR):** The IUR provides another way to assess a measure's ability to detect true variation. The IUR provides an estimate of the fraction of total variance that is due to signal (i.e., true variation in provider performance) by rescaling the *F* statistic – itself a measure of the ratio of between-groups variance to within-groups variance – using the formula (1-1/F).[2] IUR values are therefore between 0 and 1, with values close to 1 indicating a measure is reliable. An IUR greater than 0.9 is recommended for a measure to be considered reliable.
3. **Adams's rho (ρ):** This approach was developed by Adams[3] and was used for the NQF endorsed state level measure (NQF #3400) upon which our provider-level measure is based. It relies on the observed variance of the observed measure score for each as the measure of precision and produces a different reliability statistic for each provider. According to Adams, a 70% reliability demonstrates differences between providers, and a 90% reliability represents statistically significant differences between providers.

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

Performance Score Reliability Testing

Signal-to-noise analysis: Results of the signal-to-noise analysis are presented in **Table 4**. The significant F statistic ($F = 74.49$, $p < 0.001$) indicates a significant difference in performance between providers. Eta-squared and omega-squared values of 0.50 and 0.49, respectively, indicate a large effect size and thus that the differences between provider scores are large.

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

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

Table 4: Results of Reliability Testing Based on Signal-to-noise analysis

Measures	F	η2	ω2
All Providers	74.49*	0.50*	0.49*
Individual Clinicians	71.17*	0.58*	0.57*
Hospitals/Facilities/Agencies	84.84*	0.44*	0.44*

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

Intra-unit reliability (IUR): The calculated IUR was 0.99 for the combined provider sample, the individual clinicians and the hospitals/facilities/agencies. This exceeds the threshold recommended for determining acceptable reliability of 0.9.

Adams's rho (ρ): The mean calculated Adams's ρ was 0.95 for all the samples, which exceeded the 0.7 threshold recommended to indicate acceptable reliability. The median ρ was 0.96-0.97, the standard deviation was 0.05 and the interquartile range was 0.07. (See distribution of ρ in **Table** below.) The minimum observed value was 0.76, thus, 100% of providers had a ρ greater than or equal to the 0.7. threshold of acceptability.

Table 5: Results of Reliability Testing Based on Adam's rho (*ρ*)

Measures	n	Mean	SD	Min	10th	25th	50th	75th	90th	Max
All Providers	9,398	0.95	0.05	0.81	0.87	0.92	0.97	0.99	1	1
Individual Clinicians	5344	0.95	0.05	0.83	0.88	0.92	0.96	0.99	1	1
Hospitals/Facilities/Agencies	4054	0.95	0.05	0.76	0.86	0.92	0.97	0.99	1	1

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

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

2b1. VALIDITY TESTING

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

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

☐ **Performance measure score**

☐ **Empirical validity testing**

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

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

Convergent validity: Convergent validity is established by empirically showing that measures that are conceptually related are statistically correlated to one another. The *Prescription or administration of pharmacotherapy for (OUD)* measure (the "MOUD" measure) and the *Continuity of Care After Inpatient or Residential SUD Treatment at the Provider Level* (the "SUD Follow-Up" measure) are conceptually related. Both are best practices for addiction treatment. The SUD Follow-Up measure captures the percentage of Medicaid discharges, ages 18 to 64, being treated for a substance use disorder (SUD) from an inpatient or residential provider that received SUD follow-up treatment within 7 or 30 days after discharge. SUD treatment is defined to include outpatient, intensive outpatient, or partial hospitalization visits; telehealth encounters; SUD medication fills or administrations; or residential treatment (after an inpatient discharge). Two rates are reported: continuity within 7 and 30 days after discharge.

To empirically assess the relationship between the two measures, we conducted correlation analysis between the MOUD measure and the SUD Follow-up measure at 7-day follow-up and at 30-day follow-up . This analysis was limited to providers who were captured in both measures (i.e., hospitals and inpatient residential providers). We calculated the Pearson Product Moment Correlation Coefficient (r), which measures the strength of the association between the two measures. Looking at absolute values, a coefficient value of $r < 0.3$ indicates weak strength, $0.30 \leq r < 0.5$ indicates moderate strength, and $r \geq 0.50$ indicates a strong relationship.[4]

As a further test of convergent validity, we evaluated whether the *Prescription or administration of pharmacotherapy for opioid use disorder (OUD)* (the "MOUD" measure) was correlated with a measure of the percentage of that provider's patients who were hospitalized or had an emergency department visit with a SUD principal diagnosis or an overdose principal diagnosis within 30 days after an attributable encounter with the provider. An attributable encounter was defined as either two visits to the provider no more than 14 days apart or discharge from an inpatient/residential stay lasting at least 14 days.

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

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

The MOUD measure was negatively and significantly correlated with the percent of a provider's patients that were hospitalized or had an emergency department visit for a substance use disorder or opioid use disorder diagnosis within 30 days of an encounter with the provider r = -0.32 ($p < 0.001$) for all providers, r = -0.35 for individual clinicians, and r = -0.28 for hospitals/facilities/agencies.

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?*)

[4] Cohen, J. (2013) Statistical power analysis for the behavioral sciences. Academic Press. Barch, D.H. (2019) Statistics for Everybody. Dubuque: Kendal Hunt Publishing.

Convergent testing demonstrated that the MOUD measure is significantly correlated with two other measures of quality, follow-up after discharge and hospitalizations or emergency department visits for a substance use disorder or overdose following 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*)

Rationale for exclusions provided under section 2b2.3.

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

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

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

Individuals under 18 are excluded. Rationale: There is limited evidence regarding the efficacy of MOUD for this population.

Individuals over 64 are excluded: Rationale: Most individuals over age 64 are covered under Medicare. Services covered by Medicare would not be capture in the Medicaid claims data and therefore follow-up treatment would be missed. And the measures would not be valid.

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

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

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

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

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

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

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

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

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

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

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

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

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

If stratified, skip to 2b3.9

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

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

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

2b3.9. Results of Risk Stratification Analysis:

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

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

_____

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

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

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

As presented in section 2a2.3, the *F* statistic ($F = 74.49$, $p < 0.001$) indicates a statistically significant difference in performance between providers. Eta-squared and omega-squared values of 0.50 and 0.49, respectively, indicate a large effect size and thus that the differences between provider scores are meaningful. The calculated IUR was 0.99, and the mean calculated Adams's *ρ* was 0.95, which both exceeded the 0.9 and 0.7 thresholds respectively further demonstrate acceptable reliability.

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

The *F*-statistic indicates that the measure scores are significantly different while subsequent statistics indicate a large effect size, empirically substantiating that the measure can discern clear differences in performance between providers.

_____

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

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

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

set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

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

Not applicable.

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

Not applicable.

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

Not applicable.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

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

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

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

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in 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).

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

Attachment:

3c. Data Collection Strategy

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

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

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

Not applicable.

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

None.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
*	Quality Improvement (external benchmarking to organizations) New York, Massachusetts and West Virginia Medicaid https://www.treatmentatlas.org/ Quality Improvement (Internal to the specific organization) New York, Massachusetts and West Virginia Medicaid https://www.treatmentatlas.org/

*cell intentionally left blank

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

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

- Name of program and sponsor: New York Office of Addiction Supports and Services, Shatterproof ATLAS, a related measure is being used as part of the CMS Medicaid Adult and Home Core Sets Program.
- Purpose: Quality Improvement
- Geographic area and number and percentage of accountable entities and patients included: New York state (approximately 274 addiction treatment facilities), Shatterproof ATLAS (approximately 400 addiction treatment providers across 4 states, New York, Massachusetts, Delaware, and West Virginia).
- Level of measurement and setting: Provider of addiction treatment.

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

Not applicable.

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

Not applicable.

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

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

All Medicaid participating substance use disorder specialty facilities in New York, New York, Massachusetts, West Virginia , and Delaware (approximately 400 facilities).

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

The data are presented in a portal only accessible to providers, state policymakers, and providers. Providers are offered technical assistance material and training to help expand access to MOUD

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

Describe how feedback was obtained.

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

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

Not applicable.

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

Not applicable.

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

Not applicable.

Improvement

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

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

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

Despite the clear benefits of MOUD, many practitioners do not offer them to their patients with OUD. In 2018, only 40% of specialty addiction treatment facilities offered medications to treat opioid use disorder (Mark et al., 2020) and many regions of the county lack adequate supply of buprenorphine waived professionals (Abraham et al., 2020, Andria et al., 2020). A 2014 American Society of Addiction Medicine expert panel recommended that the Use of Pharmacotherapy Measure (NQF #3400) be created as a measure at the clinician level (ASAM, 2014). They note that individual providers use measures for internal quality improvement and to monitor their practices (ASAM, 2014). NQF endorsed Use of Pharmacotherapy (NQF #3400) – a measure of percentage of Medicaid beneficiaries ages 18 to 64 with an OUD who filled a prescription for or were administered or ordered an FDA-approved medication for opioid use disorder (MOUD)

during the measure year - at the health plan/Medicaid program level. However, this measure is not useful for individual providers who would like to evaluate their use of OUD medications in their patient population because it is defined at the health plan/Medicaid program level. Developing a measure at the provider/clinician level will help to identify opportunities to improve OUD treatment.

The results of the testing of this measure indicate significant opportunities for improvement. The mean provider-level score was 38.4% with provider-level scores ranging from 0% to 100%.

ADD NEW YORK

4b2. Unintended Consequences

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

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

This measure has not been implemented yet. There were no unexpected findings identified during testing of this measure.

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

This is a new measure that has not been implemented yet. No unexpected benefits were observed during testing.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

3175 : Continuity of Pharmacotherapy for Opioid Use Disorder

3400 : Use of Pharmacotherapy for Opioid Use Disorder (OUD)

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?

Yes

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

The measure is harmonized with NQF3400: Use of Pharmacotherapy for Opioid Use Disorder (OUD). The same OUD code and pharmacotherapy codes are included in both. The differences between NQF 3400 and this measure (Prescription or administration of pharmacotherapy to treat OUD), is that this measure is meant to be used at the provider level. Therefore, this measure has processes to identify providers and attribute patients with OUD to them. The measure is harmonized with NQF3400: Use of Pharmacotherapy for Opioid Use Disorder (OUD). The same OUD code and pharmacotherapy codes are included in both. The differences between NQF 3400 and this measure (Prescription or administration of pharmacotherapy to treat OUD), is that this measure is meant to be used at the provider level. Therefore, this measure has processes to identify providers and attribute patients with OUD to them.

5b. Competing Measures

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

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Not applicable.

Appendix

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

Attachment Attachment: MOUD_Receipt_Attribution_Appendix_10_29_2020.docx

Contact Information

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

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

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

Co.4 Point of Contact: Tami, Mark, tmark@rti.org, 301-636-2410-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

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

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

Expert panel members included the following:

Jennifer B. Atkins, MBA

Vice President, Network Solutions, Blue Cross Blue Shield Association

Ellen Bouchery, MS

Principal Program Analyst, Mathematica Policy Research

Teresita Camacho-Gonsalves, PhD, MA

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

Vitka Eisen, EdD, MSW

President & CEO, HealthRight 360

Joseph Lee, MD

Medical Director, Hazelden Betty Ford Foundation Youth Continuum

Miriam Komaromy, MD, FACP, DFASAM

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

Tami Mark, PhD, MBA

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

Tiffany McCaslin, MPP

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

Thomas McLellan, PhD

Founder, Treatment Research Institute

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

Executive Medical Director, UnityPoint Health Illinois Institute for Addiction Recovery

Douglas Nemecek, MD, MBA

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

Andre Ostrovsky, MD

Chief Executive Officer, Concerted Care Group

Justin Luke Riley, MBA

President & CEO, Young People in Recovery

Patricia Santora, PhD

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

Sarah Wattenberg, MSW

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

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2020

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

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

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

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

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

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

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

Ad.7 Disclaimers: The performance measure, Prescription or administration of pharmacotherapy to treat opioid use disorder (OUD), is not a clinical guideline and does not establish a standard of medical care and has not been tested for all potential applications.

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

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