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

Brief Measure Information

NQF #: 3492

Corresponding Measures:

De.2. Measure Title: Acute Care Use Due to Opioid Overdose

Co.1.1. Measure Steward: Centers for Medicare & Medicaid Services (CMS)

De.3. Brief Description of Measure: This is a population measure that indicates the rate of emergency department visits for opioid overdose events in a specified geographic region using ICD-10 diagnosis codes from claims. The outcome is defined as the incidence of overdose events per 1,000 person-years among Medicare beneficiaries greater than 18 years of age residing in the specified geographic region. The measure has been tested for use at both the county and state levels.

1b.1. Developer Rationale: The goal of this measure is to accurately quantify emergency department use due to opioid overdose among Medicare beneficiaries. Accurate and timely measurement of opioid overdose can serve several purposes: it can provide stakeholders with a tool to measure the burden of opioid overdose within a community, it enables comparison among geographies, and it allows for tracking trends and improvement over time within entities. Measure results could be used to focus resources on communities most in need and to encourage investment in reducing opioid overdose, in addition to incentivizing innovation and systems improvement.

S.4. Numerator Statement: The numerator is comprised of incident outcome events, defined as opioid overdoses that result in emergency department use, within the population residing in a specific geography.

S.6. Denominator Statement: The denominator consists of all enrolled Medicare Fee-For-Service (FFS) beneficiaries with Parts A or B, aged 18 and older residing in a measured geography (either a county or a state) during a one-year period.

S.8. Denominator Exclusions: None

De.1. Measure Type: Outcome

S.17. Data Source: Claims

S.20. Level of Analysis: Population : Community, County or City, Population : Regional and State

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? N/A

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary

Logic model: 4 antecedents to opioid use disorder (OUD) and overdose: 1. Availability of prescription opioids, 2. Exposure to dangerous combination prescriptions (e.g., benzodiazepines and opioids), 3. Access to medication-assisted treatments (e.g., buprenorphine, methadone), 4. Access to OUD treatment more generally (e.g., within the justice system).

Evidence for a connection between outcome and at least one healthcare process:

Developers present a short narrative with four points:

1. Increased hydrocodone and oxycodone or high dose opioid prescribing in 1995-2004 correlated with risk of overdose (Wisniewski et al., 2008; Gwira et al., 2014; Khan et al., 2019)
2. Prescription drug monitoring programs (PDMP) "may" have favorable impacts (Gwira et al., 2014)
3. Provider-level interventions such as education, feedback, and "changing default settings for electronic prescribing may reduce opioid prescriptions (Brookings "research roundup", 2018)
4. Buprenorphine "may" reduce opioid use (Wen et al., 2018)

Additional information provided by the developer on 12/16/2019:

The measure developer notes that the initial evidence review deliberately focused on a narrow question of whether health care-related processes are associated with *population-level emergency department use for opioid overdose*. The most consistent evidence suggests that opioid prescribing, a health care process, is associated with population-level rates of ED visits for opioid-related conditions. However, the developer also recognizes that there is a much broader evidence base linking health care interventions and processes to opioid overdose. Accordingly, the measure developer has provided a supplementary summary of the evidence which broadly describes the relationship between health care interventions, processes, and the outcome of opioid overdose.

Supplemental references and a summary provided by the developer further support that reducing high-risk prescribing and providing MAT can reduce opioid misuse and risk of overdose.

Questions for the Committee:

- Is the evidentiary presentation sufficient to causally link this measure to healthcare processes that are discernable separate from the effects of non-healthcare factors? (This was a substantive concern of the Scientific Methods Panel (SMP), even as they were advised to defer to the Standing Committee regarding this determination. Note that the SMP did not review the evidentiary presentation because such material is outside of their explicit scope).

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

Health outcome (box 1)→ the relationship between the measured health outcome and one healthcare action is demonstrated by empirical data (box 2)→ Yes → PASS

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.

Below are state-level results for each year, in ED visits per 1000 person-years (mean (SD); Range [min, max]):
2017: 1.11 (0.36); Range 0.53-1.80
2018: 0.92 (0.33); Range 0.42-1.71

Below are county-level results for each year (mean (SD); Range[Min-max]):
2017: 1.55 (1.11); Range 0.59-6.22
2018: 1.44 (1.23); Range 0.48-6.44

Narrative seems useful. For example, in 2014, Massachusetts had a rate of 450 visits per 100,000 population while Iowa had a rate of 45 per 100,000 (Weiss et al., 2017).

Disparities
Social determinants are evident, they show by a brief citation of these:

1. Brady JE, Giglio R, Keyes KM, DiMaggio C, Li G. Risk markers for fatal and non-fatal prescription drug overdose: a meta-analysis. Inj Epidemiol. 2017;4(1):24.
2. Dasgupta N, Beletsky L, Ciccarone D. Opioid Crisis: No Easy Fix to its Social and Economic Determinants. Am J Public Health. 2018 Feb; 108(2) 182-186.
3. Abraham AJ, Andrews CM, Yingling ME, Shannon J. Geographic Disparities in Availability of Opioid Use Disorder Treatment for Medicaid Enrollees.

Questions for the Committee:

- None

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

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

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."

- This is a population measure, so I am not sure if this question is relevant.
- This only measures in medicare population and does not differentiates the substance.

- Evidence exists that opioid prescribing is correlated with risk of overdose; prescription drug monitoring programs are beginning to evidence positive impact on prescribing behavior; provider interventions may reduce opioid prescriptions.
- The intervening effect of SDOH, reporting/referral biases, and differences in protocols of care, undermine the integrity of this measure.

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

- County level measurement is important with respect to opioid overdose. It is required by SAMHSA in order for states to target resources to counties hardest hit and to track improvements over time with greater geographic specificity and granularity.
- Moderate
- Performance gap exists for this measure between states.
- Clearly, we have an opioid abuse problem. We also have ERs with a substantial burden of care for ODs. Unfortunately, this measure does not account for the many covariates that could reasonably expect to influence this outcome. Disparities in care also may be accentuated or masked, and there is not a fulsome evaluation of these factors.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

2c. For composite measures: empirical analysis support composite approach

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Methods Panel Review (Combined)

Scientific Methods Panel (SMP) Subgroup Votes
Reliability: H-1; M-2; L-1; I-0 (Pass)
Validity: H-0; M-1; L-2; I-1 (Did Not Pass)

Methods Panel Evaluation Summary:
In their preliminary analyses, subgroup members of the SMP responsible for review of this measure did not reach consensus on the reliability or validity. Concerns that led to this ambivalence were largely about the specification of the measure in Medicare Part A and B individuals exclusively, rather than broader assessment of all-payer populations. Validity testing specifically was identified as a concern because there was some reliance on an advisory committee composed only of personnel from the developer's home institution, and there was a perceived dearth of details about the empirical validity testing that was conducted. Moreover, the developers did not do risk adjustment to generate their measure even as they cited studies that sociodemographic factors exogenous to the healthcare system influence the rates being measured.

During the in-person meeting this measure was discussed further. Concern persisted about the narrow data scope (Medicare A and B enrollees), the notion that interventions could be evenly disseminated at the geographic levels specified, the face validity composition and description, and the absence of exclusions (i.e., for hospice) composing this measure.

Developers provided score-level empirical validity testing results showing that this emergency department overdose measure correlated reasonably with two separate measures of opioid-related death in one case, or hospitalization in a second case. However, subgroup members noted that these comparisons were not limited to the same Medicare population and the same type of events. Some SMP members, although not all, expressed concern regarding the lack of risk adjustment for the measure, particularly adjustment for social demographic factors like the ones described in their gaps discussion. At least one SMP member suggested this measure may be one in which presenting both unadjusted and adjusted results could be important and informative.

Following in-person deliberations the subcommittee re-voted on this measure and passed it on reliability but failed it on validity. This measure is eligible for a re-vote by the Standing Committee as it was pulled by a Standing Committee member.

Additional information provided by the developer on 12/16/2019:

Developers recognize that specifying and testing the measure in the Medicare population may limit generalizability. However, the measure focuses on Medicare beneficiaries for two reasons. First, the measure was developed for use in the Maryland Total Cost of Care model, which applies to Medicare beneficiaries. Second, Medicare claims include enrollment information, which is needed to calculate the denominator. Comparable all-payer data sets, such as the National Emergency Department Sample, include claims, but no direct way to calculate a denominator that reflects the size of the measured population. These points were discussed during the SMP in-person meeting.

More broadly, with respect to validity, developers note that they performed empirical testing as the primary basis for validity assessment. Empirical testing compared the proposed measure to two independent measures (state-level rates of fatal overdose and state-level rates of opioid-related ED visits and hospitalizations) and showed strong correlations between the proposed measure and the two comparison measures. As is common in empiric testing, measured populations were not identical between the proposed measure and comparison measures. However, measure developers assert that observing strong correlations despite differences in the measured populations actually bolsters the case for validity, and suggests that the measures are capturing the

same underlying construct, despite differences in populations. During SMP discussion, SMP members did not have additional objections to empiric validity testing methods or results. Measure developers note that face validity assessment was performed to supplement empirical validity testing, not as the basis for the assessment of validity.

Questions for the Committee regarding reliability:

- None

Questions for the Committee regarding validity:

- Is risk-adjustment of this measure warranted in some form to justify its inclusion or exclusion in the measure?
- Is this measure specified at a level that makes it sensitive to healthcare program changes, or are state- and county-level analytics too broad?
- Does limiting this measure to Medicare A&B enrollees compromise its overall validity beyond acceptable?

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

Details about the reliability and validity of this measure appear below in a section entitled "Scientific Acceptability: Preliminary Analysis Form"

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?

- The data elements appear to clearly defined. However, I was not able to discern from the Submission uniform criteria in the measure for how hospital emergency departments identify an opioid overdose.
- Low
- Measure is reliable for the intended population.
- I worry about the reliability based on differences in coding and attribution.

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

- Only that it appears to measure only FFS Medicare and enrollees in Medicare Advantage plans.
- YEs
- none.
- Yes, I am concerned about limited testing in a limited population which is probably substantially different than the population more broadly.

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

- No, other than the fact that SAMHSA requires states to report all emergency room overdoses across all payors for the purpose of allocating resources.

- yes
- Focuses exclusively on Medicare populations.
- Yes, this is a fatal flaw. Face validity is insufficient in this instance.

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?

- This is a population measure and does not assess quality. It has only limited validity in assessing population health as it is restricted to FFS Medicare and does not appear to be unduplicated, i.e. patients with repeated relapses within the reported period of time.
- missing data
- Missing data not addressed.
- Lots of threats to validity, no adjustments for known confounders. Not ready for prime time IMHO

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?

- This measure has significant and important exclusions as it is restricted to FFS Medicare. As a population measure it is to be used assess overall public health performance of a community. However, only measuring Medicare enrollees is way too limited in scope.
- DNA
- No risk adjustment performed. This measure would be impacted by social risk factors..
- Ditto, as above.

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.

Medicare claims, no major challenges. Electronic, no fees associated with use of the measure.

Questions for the Committee:

- None

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?

- As far as I could tell this measure does to utilize EHRs. Further, 42 CFR Part 2 is a significant barrier to all data collection in opioid treatment.
- ED diagnosis frequently inaccurate
- Medicare Claims. No issues.
- Feasible, although flawed

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

This measure would be used to track Maryland's progress on reducing opioid overdoses under the Maryland Total Cost of Care Model.

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

Additional Feedback: N/A

- Is the use discourse sufficient for this new measure?

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

RATIONALE: Additional details about the use of this measure and feedback from those being measured will be important during maintenance review.

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 N/A

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 N/A

Potential harms: N/A

Additional Feedback: N/A

Questions for the Committee:

- None

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

RATIONALE: New measure. This is not a "must-pass" criterion. At present, there is little information presented.

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?

- SAMHSA requires states to collect and submit emergency room opioid overdose data for the purpose of allocating federal funds. All states have been doing this since 2016.
- Does not indicate prescription from street opiates
- No issues.
- Little or no data presented. Not must pass, correct? It would fail on this aspect if a must pass.

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.

- This data is only usable for the purpose of improving population health if it is collected across all payors and the uninsured.
- not demonstrated
- Would be useful for Medicare, but limited utility elsewhere.
- As above.

Criterion 5: Related and Competing Measures

Related or competing measures
None reported. The developers were advised to consider related measure 3501e (Hospital Harm- Opioid Related Adverse Events).

Harmonization
None reported (see suggestion above).

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. SAMHSA requires states to collect this data. The restriction to FFS Medicare is not harmonized with this federal requirement.
- none known
- --
- ?don't know

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/23/2020

- I support the adoption of this measure for several reasons:

  *Nonfatal overdose is a serious event that is highly predictive of fatal overdose

  *Population-based measures (meaning a geographic population) reflect community health.

  *There are important steps that the clinical community can take to reduce the risk of nonfatal overdose in a geographic population, such as improve access to medications for opioid use disorder, which are associated with declines in death rates of 50% or more.

  Other related measures could include fatal overdoses in a geographic population.
- I support the adoption of this measure.
- The proposed measure has a clearly and appropriately defined numerator and denominator. The rate of ED visits due to opioid overdose events is an important indicator of prevalence and severity of opioid use disorder in a given geographic area, and thus can provide useful guidance to officials

working to identify public health priorities, allocate public-health and clinical resources, and gauge the effectiveness of interventions targeting opioid use. For these reasons, I support the measure as submitted. Thanks for your consideration.

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3492

Measure Title: Acute Care Use Due to Opioid Overdose

Type of measure:

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

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

Data Source:

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

Level of Analysis:

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

2. **Briefly summarize any concerns about the measure specifications.**

   Panel Member #1: No concerns regarding clarity of specifications. Two observations:

   [1] One issue regarding the specification is a contraction in the MIF, S.7. It states "The denominator reflects the size of the population in which overdose events occur".

   However, later in S.7 it states "Eligible beneficiaries are assigned to geographies based on place of residence. Thus, individuals contribute to the denominator and the numerator based on residence rather than where the event took place.".

   [2] I'm unclear whether the denominator contains Medicare Advantage & FFS while the numerator only includes Medicare FFS. Would be preferable to know. If so, rates (per the measure) are lower than they are in reality. However, given the limitation applies to all counties and states, no bias is introduced to a given state / county.

Panel Member #2: For some reason, the specifications were difficult to follow and confusing. I would recommend putting all the specifications in one document and clearly label the elements. Also, not sure if there is a time period for eligibility e.g., does the beneficiary have to be enrolled for XX period of time to count? What about gaps in enrollment? Are they including duel eligible in the measure? It seems like this is a surveillance measure.

Panel Member #5: I am concerned about the denominator (enrolled Medicare FFS >= 18 yrs old). How do these numbers compare with the # of residents >= 18 years old in each county/state? Would total # of ER visits during the time period be a better denominator what is proposed? Measure would be a ratio of # visits due to opioid overdose / total number of visits. Medicare FFS >=18 enrolled would under report (lower rates) for states with higher older (mandatory) Medicare FFS enrollment (i.e., >=65) and vice versa for states with lower older populations.

Panel Member #6: The measure specs are clear. I had to contemplate the title for this one. I wanted to interpret the title as % of opioid overdoses that resulted in ED utilization, but that would be a different measure and require different data sources (somehow get the number of all opioid overdoses that do not come to and ED, and calculate). This measure does NOT ask that question. Instead, it says if you have an opioid overdose AND show up in an ED, what is that rate in all of the Medicare population in a specific area.

RELIABILITY: TESTING

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

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☐ **No**

   Panel Member #2: Testing was done, but none of the specifics were reported.
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☒ **No**

   Panel Member #1: NA – score level reliability test conducted

   Panel Member #5: (NA = X)
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: No concerns. Tests seems reasonable given the measure.

   "it is meaningful to consider the notion of "unit" reliability, that is, the reliability with which individual units (here, states or counties) are measured." [p7]

   "also performed split sample reliability testing. For spilt sample reliability testing, we randomly divided the sample into two parts and compared measure outcomes in each half of the data using a correlation coefficient." [p8]

   Panel Member #2: Developer did not provide a lot of detail around the testing methods used and didn't provide any data other than ranges and averages, which makes an assessment of the methods used difficult.

   Panel Member #3: Split sample testing

   Panel Member #4: The methods used for reliability testing seem to be appropriate.

   Panel Member #5: Split half correlation is appropriate.

Panel Member #6: Split-sample testing was utilized for reliability testing at the relevant geographic level resulting in kappa statistic of 0.92 to 0.98 at the state level for years 2009-2012 and 0.70 to 0.87 at the county level for the same years. These are stated to be the most recent years of testing.

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

   Submission document: Testing attachment, section 2a2.3

   Panel Member #1: Test results demonstrate high reliability of the measure at state and county level.

   "Among the 25 states evaluated … Adams reliability ranged from 0.92-0.99 with a mean of 0.98. Among counties in Maryland, Adams reliability ranged from 0.60-0.99 with a mean of 0.89. Among counties, only a single county had a reliability score below 0.7. " [p9]

   "Split sample reliability testing indicated a high correlation (r=0.94) between split samples at the state level. At the county level, measure results between the split samples had a correlation coefficient of 0.87." [p9]

   Panel Member #2: Developer did not provide a lot of detail around the testing methods used and didn't provide any data other than ranges and averages, which makes an assessment of the results difficult.

   Panel Member #3: Reliability score >0.9 -- strong

   Panel Member #4: Results of reliability testing indicated that the measure is highly reliable at the state level and generally reliable at the county level.

   Panel Member #5: Reported correlation values (mean = 0.94 across 25 states) seems unreasonably high if the split halves are truly random selections.

   Panel Member #6: Taken at face value, the reliability is moderate. The data is relatively old, apparently has not been reproduced, and have not been presented in a way to assess reliability at an individual county level, nor at an individual state level.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☒ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and 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)

Panel Member #6: Without more details, it is hard to assess reliability adequately. We have summary statements about reliability at the state level and county level, but have no data to assess the actual data from which this was derived.

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

Panel Member #1: Unit level reliability, using Adam's method, was generally high with results ranging from 0.92 to 0.99 in regard to states. Regarding counties, mean result was 0.89 where only one county was below 0.70 (i.e. (0.60).

Panel Member #2: The developer did not provide enough data to accurately assess the measure's reliability. See comments above.

Panel Member #3: Fairly simple claims based methodology that has been previously tested on more local level

Panel Member #4: Previously published data indicated that the enrollment data elements and codes used in ED claims are generally reliable. Data testing at the measure score level showed high levels of reliability using appropriate tests.

Panel Member #5: The reported split half correlation seem unreasonably high for a randomized split of 25 different data sets. Did the developers compare a county's 2017 results to that county's 2018 results? I could believe that those correlations would be in the range of 0.94.

Panel Member #6: See above.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member #1: NA - No exclusion

Panel Member #3: None

Panel Member #4: Focusing only on Medicare FFS patients when Medicare Advantage penetration rates vary by both state and county is a threat to validity, as is the absence of a distinction between patients under 65 who are eligible by way of disability and those over 65 eligible by way of age. The problems seem manageable, though, since correlations with similar measures in other populations not subject to these problems are reasonably high.

Panel Member #5: No exclusions reported. Agree.

Panel Member #6: There are no exclusions.

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

Submission document: Testing attachment, section 2b4.

Panel Member #1: No concerns. We see a fair amount of statistically tested differences in states & counties. [p17]

Panel Member #2: No exclusions.

The developer provided some data about the ability to identify meaningful difference in performance, but there should have been more information about the methods used, e.g., statistical tests and additional data to assess their claims of meaningful differences across counties and states.

Panel Member #3: None

Panel Member #4: The measure developers make no attempt to define or identify meaningful differences. Identifying meaningful differences is not a stated purpose of the measure.

Panel Member #5: Two issues:

1. See concerns about proper denominator stated previously.

2. While I agree that this metric may be a useful monitoring metric, the reason for higher or lower rates by county or state is not identified—and methods for improving rates and who (county, state, Providers within these units) are not directly obvious. Perhaps if State A implements a statewide intervention program and produces a lower rate than State B that does not, that could indicate the potential value of such an intervention program—assuming that all other key influencing variables ae the same between these two states. However, comparing rates for WY and FL would be virtually impossible given geography, climate, and population—to name just a few differences.

Panel Member #6: Data is presented that meaningful difference in performance with rates summarized at the state level with confidence intervals given. 12 states were below the mean, 3 at the mean, and 10 above the mean. Data is also presented for 25 counties from 2017 with 2 above the mean and 9 below the mean for counties.

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.
Panel Member #1: NA – only 1 data source employed.
Panel Member #2: Not applicable.
Panel Member #3: N/A
Panel Member #5: No comment.

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

Submission document: Testing attachment, section 2b6.

Panel Member #1: NA – Testing form has a blank in regard to 2b6 & "NA" is stated in the 3 subsequent questions. Thus, assume no missing data.

Panel Member #2: The developer did not address the missing data issue.

Panel Member #3: Focuses on claims items that are frequently audited

Panel Member #4: No data appear to be missing or unavailable.

Panel Member #6: Missing data is not addressed.

16. **Risk Adjustment 2b3**

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

Panel Member #2: The developer provided conceptual rationales for not risk adjusting the measure. I would have liked to see some analysis of the data to support their rationale.

Panel Member #5: Although rationale for not risk adjusting is provided, I believe that there are some obvious relationships between opioid use and sociodemographic variables such as age (e.g., opioid overdose is seldom reported in the senior citizen population), economic conditions (e.g., opioid usage

is often associated with poorer economic conditions), among other variables. I would suggest looking at these rather than race.

16c. **Social risk adjustment:**

Panel Member #2: Not applicable.

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

Panel Member #5: See previous comment

16d.**Risk adjustment summary:**

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

Panel Member #1: NA-not risk adjusted

Panel Member #5: (NA = X)

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

Panel Member #1: NA-not risk adjusted

Panel Member #5: (NA = X)

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

Panel Member #1: NA-not risk adjusted

Panel Member #5: (NA = X)

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

Panel Member #1: NA-not risk adjusted

Panel Member #5: (NA = X)

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

Panel Member #1: NA-not risk adjusted

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

Panel Member #1: NA – Measure is not risk adjusted. One question / concern:

The MIF in S.7 states the following:
"Identifying emergency department visits requires information from both inpatient and outpatient claims which are covered by Medicare Parts A and B respectively. In order to be maximally inclusive, the measure includes all beneficiaries with either Part A or B, rather than requiring that beneficiaries have Parts A and B. Limiting the measure to beneficiaries who have Parts A and B would exclude individuals with observable outcome events. For example, beneficiaries with Part A would have observable outcome events if they are admitted to the hospital for an opioid overdose while those with Part B would have an observable outcome event if they were seen only in the emergency department. Although this approach may miss some outcome events for beneficiaries with only Parts A or B, it allows the measure to be maximally inclusive of both the measured population and potential outcome events." [p5]
Given the above stated specification, the implication is that regions (i.e. counties, states) that have higher rates of that beneficiaries that have both Parts A and B will have higher rates than regions with beneficiaries that have Part A only, Part B only. Ideally, the measure steward would have performed appropriate testing to ensure this introduced no substantial bias. However, I do not see this addressed in the MIF nor Testing form.

Panel Member #3: The purpose here is to provide regional data which describes an important and increasingly troublesome health issue rather than to evaluate the quality of services provided—risk adjustment would corrupt the intention of the metric.

Panel Member #4: The measure developers have chosen not to do risk adjustment, even though a number of social or economic variables have strong relationships to the "outcome" being measured. They argue that adjusting would mask important disparities. In the context of this measure, this is not a problem. There is no linkage of the measure to quality of care, and the measure does not measure quality of care. Risk adjustment is important when plans or providers or other entities are being compared with each other, and comparisons are presumed to reflect differences in quality of care. Since this is not a quality measure and no specific comparisons among entities are proposed, the absence of risk adjustment does not have the usual problems that it would have with quality measures.

Panel Member #6: The measure developer states that risk adjustment is not necessary because the risk of opioid overdose varies according to patient demographic characteristics, thus social and not biological differences. Analysis could obscure health disparities. Also, risk adjustment would allow for higher rates in some populations and would perpetuate inequities in the health of communities.

Nonetheless, analysis was performed among the 25 states and observed low or moderate correlations with white/non-Hispanic, median household income, percent in poverty, and living in rural areas.

For cost/resource use measures ONLY:

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

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

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING 2b1

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

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

    ☒ **Face validity**

    ☒ **Empirical validity testing of the measure score**

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

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

    Submission document: Testing attachment, section 2b1.2

    Panel Member #1: Regarding empirical testing: Types of tests and data sources employed are appropriate. Regarding the use of the AHRQ measure, while the data is claims (which is the same as the draft measure, the issue is the AHRQ measure apparently includes all payer cases. Where the draft measure restricts the population not Medicare cases only.

    Regarding face validity testing: Undesirable for Yale to solely engage Yale clinicians on the panel. Clinicians from various setting would have been more desirable where none were from Yale.

    "Measure Score Validity - Empirical Testing Through Comparison to Independent Data Sources:

    …we compared the proposed measure to two other sources: a claims-based measure of opioid-related acute care use independently developed and reported by

    [1] … AHRQ

    [2] state-level opioid overdose death rates from the [CDC]." [p10]

    "Measure Face Validity:

…we convened a meeting of clinicians with expertise in opioid use and emergency medicine… included four practicing emergency medicine physicians, all with research expertise in opioid use disorder and in quality measurement, and two general internists with expertise in opioid use disorder…. All members of this group were faculty members at the Yale School of Medicine. We presented information on the measure context, potential uses, and measure specifications to this expert panel. We also presented a version of this measure with a much broader outcome definition that may be more sensitive but less specific. This broader definition includes all ICD-10 codes included in the present measure. … 5 members voted on the face validity of the measure. All 5 voting members ranked the measure a 4 on a Likert scale." [p11]

Panel Member #2: A panel was convened to establish face validity, but the panel members were not specified and the methods used to vote on the measure were not discussed. The developer did not provide the details of the empirical testing.

Panel Member #3: Historical validation from the literature of the data element validity. Measure tested against NEDS database and conducted face validity testing with Yale group of experts

Panel Member #4: The methods are appropriate for an epidemiological measure that is not presumed to have anything to do with quality. Basically, "it is what it is", and face validity carries significant weight.

Panel Member #5: Methodology for new measure is appropriate.

Panel Member #6: The measure score was compared to two sources of opioid-overdose data – AHRQ and state level from the CDC. It is stated that they compared favorably but the data is not provided.

Face validity was provided by a convened TEP.

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

Submission document: Testing attachment, section 2b1.3, 2b14

Panel Member #1: Regarding empirical testing: Correlation between the draft measure and the two other sources were high: both at 0.74. Some concern regarding:

[1] no other correlation information presented, e.g. correlation for higher and lower volume regions.

[2] Draft measure rates were 0.86 – 1.84 & the AHRQ comparison yielded rates several fold greater: 2.27 – 9.6. The measure developer neglected to discuss the difference in rates. Much of the difference is likely due to AHRQ's measure inclusion of all cases where the draft measure is restricted to Medicare cases (but the measure developer was silent on whether they controlled for this population difference in 2b1.2 (above).

Regarding face validity testing: While the measure developer describes the testing they failed to share these testing results (e.g. use of a Lickert scale). Some discussion of group engaged in face validity was summarized in 2b1.2. However, 2b1.2 asks to describe the test (not the results of the test) while 2b.1.3 is to relay the testing results.

"In 2017, the rate of ED visits for opioid overdose in the proposed measure ranged from 0.86-1.84 per 1,000 person-years among the 25 states tested. In NEDS/NIS data, which is an all-payer population, rates ranged from 2.27-9.60 per 1,000 population per year. Correlation between the two measures was high (r=0.74).

Correlation between the proposed measure outcome rate and opioid overdose death rate was also high (0.74) at the state level among 25 states tested."

Panel Member #2: I need the details to adequately assess what was done, which score was compared to which, and I think that if they are measuring a Medicare population, that they should be using that population for validation rather than an all-payer database.

Panel Member #3: High level of correlation (0.74)

Panel Member #4: Results of validity testing show acceptably high levels of agreement with independent measures of very similar concepts.

Panel Member #5: A 5-member TEP seems rather small for such an important measure. I suggest redoing this analysis with at least 15 members that are drawn from a variety of geographic regions, population densities (urban/rural) , and economic strata.

Panel Member #6: Hard to assess as source data is not provided

23. **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)

    Panel Member #6: Comparison to above databases cited.

24. **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**

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

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

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

    ☐ **Low** (NOTE: Should rate LOW if you believe that there <u>are</u> threats to validity and/or relevant threats to validity were <u>not assessed OR</u> 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 <u>is required</u>; if not conducted, should rate as INSUFFICIENT.)

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

    Panel Member #1: <u>Empirical testing</u> results indicate validity, but (as noted above in Q22), concerns include:

    [1] no other correlation information presented, e.g. correlation for higher and lower volume regions.

    [2] Draft measure rates were 0.86 – 1.84 & the AHRQ comparison yielded rates several fold greater: 2.27 – 9.6. The measure developer neglected to discuss the difference in rates. Much of the difference is likely due to AHRQ's measure inclusion of all cases where the draft measure is restricted to Medicare cases (but the measure developer was silent on whether they controlled for this population difference in 2b1.2 (above).

    <u>Face validity</u>: Also (as stated in Q22), while the measure developer describes the testing they failed to share these testing results (e.g. use of a Lickert scale). Some discussion of group engaged in face validity

was summarized in 2b1.2. However, 2b1.2 asks to describe the test (not the results of the test) while 2b.1.3 is to relay the testing results.

Due to these concerns, refrained from a high score.

Panel Member #2: To my way of thinking there was not enough information provided by the measure developer to adequate assess the measure. While it appears that they did some sort of testing, I can tell what was done, which statistical tests were done, or what they were comparing. Further, I have a concern using a different population than the measure calls for to validate the measure.

Panel Member #4: The measures has reasonable validity as an epidemiological measure of rate of opioid overdose events at the state and county level. It has no validity, nor does it attempt to establish any validity, as a measure of quality of care or any other dimension of "quality".

Panel Member #5: See previous comments related to appropriate denominator for metric, lack of risk adjustment, and inadequate face validity panel size.

Panel Member #6: From the data provided, I am unable to assess either the face validity or the independent database comparative validity correlation rate.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

ADDITIONAL RECOMMENDATIONS

29. **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.**

    Panel Member #4: Measures like this should not be coming to the SMP or probably to NQF at all. Our processes and definitions and review criteria are set up to evaluate quality measures. This is not a quality measure. Although it is possible to answer most of the review questions, it should not be presumed that any "passing grade" or ultimate NQF endorsement implies that this is a measure of the quality of anything. No such standing has been established here. I would prefer that NQF not be in the business of endorsing measures like this. This is straightforward population epidemiology, and NQF resources would be overwhelmed if it took on the task of reviewing and endorsing all such possible measures.

    Panel Member #5: I know that we need measures to monitor overall progress (or lack thereof) for this problem. However, simply reporting numbers of questionable merit and no opportunity to identify leverage mechanisms to create positive change to address this problem seem counter-productive.

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

7.31.19_Maryland_nqf_evidence_attachment_v1.0.docx

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

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

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 3492

Measure Title: Emergency Department Use Due to Opioid Overdose

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: Click here to enter a date

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

Outcome

☒ Outcome: Opioid overdose events

☐ Patient-reported outcome (PRO): Click here to name the 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*): Click here to name the intermediate outcome

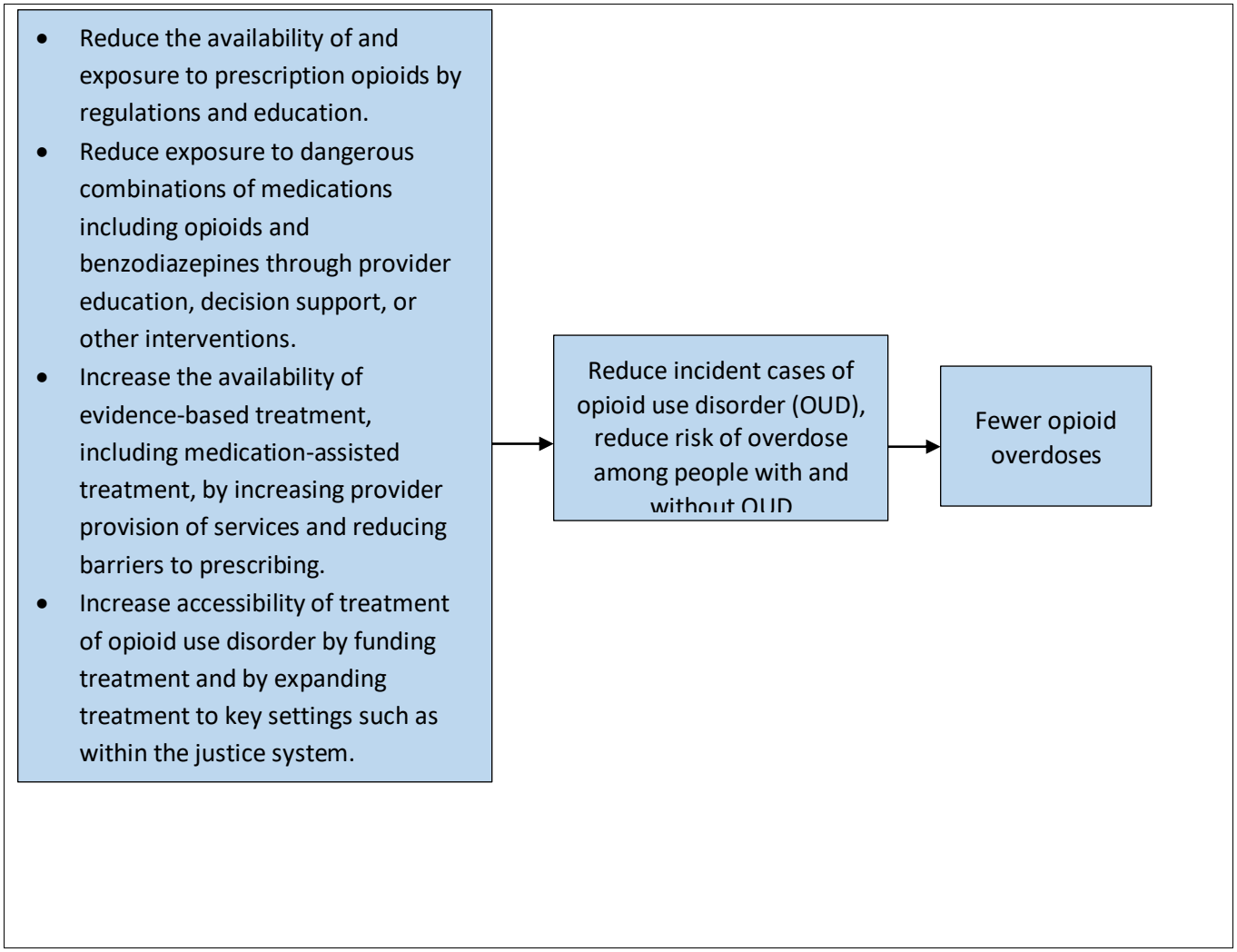
☐ Process: Click here to name what is being measured

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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.



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. This measure is not derived from patient report.

****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.

Opioid overdose is associated with a range of health care processes including opioid prescribing and availability of treatment for opioid use disorder. Some of the strongest evidence indicates that opioid prescribing is associated with opioid use disorder and overdose, and evidence supports a number of strategies to reduce opioid prescribing. A study by Wisniewski et al. characterized the relationship between national opioid prescribing and ED visits for opioid toxicity using nationally representative data from 1995 to 2004 (1).

The authors found a statistically significant correlation between hydrocodone and oxycodone prescribing rates and ED visits (Kendall's tau 0.73 and 0.87, respectively, $p<0.05$) during the years evaluated. On the individual level, receipt of a prescription for high dose opioids increases risk of overdose among individuals and receipt of any prescription increases risk among family members (2, 3).

Policy-level interventions can be effective at reducing unsafe prescribing and downstream health effects. Prescription drug monitoring programs (PDMPs), which provide prescribers with information about patients' opioid prescriptions may reduce prescribing and may result in lower rates of opioid prescribing and opioid use disorder (2). Provider-level interventions, such as education, feedback, and changing default settings for electronic prescribing may also reduce opioid prescriptions (4). Lastly, increasing access to treatment of opioid use disorder, particularly by increasing the availably of medication assisted treatment, may reduce opioid use and ultimately, opioid overdose (5).

References:

1. Wisniewski AM, Purdy CH, Blondell RD. The epidemiologic association between opioid prescribing, non-medical use, and emergency department visits. J Addict Dis. 2008;27(1):1-11.
2. Gwira Baumblatt JA, Wiedeman C, Dunn JR, Schaffner W, Paulozzi LJ, Jones TF. High-risk use by patients prescribed opioids for pain and its role in overdose deaths. JAMA Intern Med. 2014 May;174(5):796-801.
3. Khan NF, Bateman BT, Landon JE, Gagne JJ. Association of Opioid Overdose With Opioid Prescriptions to Family Members. JAMA Intern Med. 2019 Jun 24. doi: 10.1001/jamainternmed.2019.1064. [Epub ahead of print]
4. https://www.brookings.edu/blog/up-front/2018/12/07/research-roundup-what-does-the-evidence-say-about-how-to-fight-the-opioid-epidemic/, accessed 2/7/19
5. Wen H, Hockenberry J, Pollack HA. Association of Buprenorphine-Waivered Physician Supply With Buprenorphine Treatment Use and Prescription Opioid Use in Medicaid Enrollees JAMA Netw Open. 2018;1(5):e182943. doi:10.1001/jamanetworkopen.2018.2943

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

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

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.

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

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

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

Supplemental evidence provided by the developer:

Overview

Health care providers and health systems play a key role in addressing the opioid epidemic. Here, we highlight two mechanisms through which health care processes and interventions directly influence the risk of opioid overdose in a population. First, by **reducing high-risk prescribing** and, second, by **improving access to evidence-based treatment for opioid use disorder**, health care providers and health systems can reduce the risk of opioid overdose in their communities.

High Risk Prescribing

Opioid prescribing, particularly in high doses or in combination with benzodiazepines, is associated with increased risk of opioid overdose.[1, 2] Risk of overdose is elevated not only among people who receive a prescription, but also among people living in the same household, highlighting the far-reaching impact that opioid prescriptions have on risk of overdose.[3] Further, the association between opioid prescribing and adverse events has been documented not only at the individual level, but also at the national level. As opioid prescriptions in the US rose, the number of emergency department visits for opioid-related conditions also increased.[4]

Policies designed to limit high-risk opioid prescribing have had success in reducing both high risk prescribing and risk of overdose. For example, prescription drug monitoring programs (PDMPs) are state-level databases that allow prescribers and pharmacist to view all opioid prescriptions for a patient and to identify high risk prescription patterns. The implementation of mandatory PDMPs has been associated with a reduction in high-risk prescribing.[5] Use of state PDMPs is also associated with lower rates of ED visits and hospitalizations for opioid-related conditions including overdose.[6]

Other interventions aimed at reducing the supply of prescription opioids have reduced the population-level rate of opioid overdose. For example, as the opioid epidemic accelerated, the state of Florida aggressively implemented a number of interventions aimed at high-risk prescribing including closing pain clinics, revoking DEA certificates, and capping dispensing quantities. Implementation of these policies resulted both in decline of opioid prescriptions and a marked decline in the fatal overdose rate.[7, 8]

Policies and strategies intended to change physician behavior have also been successful at reducing high risk prescribing.[9] For example, use of an audit-feedback system, which alerted physicians when a patient had died of an overdose, resulted in a decline in opioid prescriptions from those physicians.[10] Likewise, behavioral interventions like changing default settings on electronic prescriptions to a lower quantity, may also be effective.[5] Given the link between opioid supply and overdose rates, these kinds of prescriber-level interventions may ultimately reduce risk of opioid overdose in a population.

Provision of Medication-Assisted Treatment (MAT)

In addition to their role as opioid prescribers, clinicians also play a key role in treating opioid use disorder. Clinicians can directly reduce the risk of opioid overdose among people with opioid use disorder by helping patients access medication assisted treatment (MAT). Medication assisted treatment refers to the use of medications (primarily methadone, buprenorphine, and naltrexone), along with counseling and other therapies, to treat opioid use disorder. A number of randomized controlled trials have convincingly demonstrated that MAT helps people with OUD to stay in treatment and avoid relapse.[11] High quality

observational studies consistently demonstrate a lower risk of fatal overdose after the initiation of MAT.[12] Given this body of evidence, MAT has become a cornerstone of the treatment of opioid use disorder.

Expanding access to MAT has been proposed as a key strategy in combating the opioid epidemic and preventing opioid overdose in a population.[13] Evidence-based implementation strategies include reducing barriers to access (for example by eliminating prior authorization), providing MAT in novel settings like the emergency department, primary care, or in criminal justice settings, and increasing the supply of prescribers in a population.[14, 15] Indeed, increasing the supply of providers who can prescribe MAT is associated with lower rates of population-level opioid prescription, which is likely to translate into lower risk of overdose in the population.[16]

In summary, health care providers play a key role in the opioid epidemic. Prescriber actions can both contribute to risk of overdose and mitigate risk. Population-level strategies that target opioid prescribing and treatment of opioid use disorder can reduce opioid misuse and reduce the risk of overdose.

References

1. Sun, E.C., et al., *Association between concurrent use of prescription opioids and benzodiazepines and overdose: retrospective analysis.* BMJ, 2017. **356**: p. j760.
2. Dowell, D., T.M. Haegerich, and R. Chou, *CDC Guideline for Prescribing Opioids for Chronic Pain—United States, 2016.* JAMA, 2016. **315**(15): p. 1624-1645.
3. Khan, N.F., et al., *Association of Opioid Overdose With Opioid Prescriptions to Family Members.* JAMA Internal Medicine, 2019. **179**(9): p. 1186-1192.
4. Wisniewski, A., C. Purdy, and R. Blondell, *The Epidemiologic Association Between Opioid Prescribing, Non-Medical Use, and Emergency Department Visits.* Journal of addictive diseases, 2008. **27**: p. 1-11.
5. Doleac, J., A. Mukherjee, and M. Schnell. *Research roundup: What does the evidence say about how to fight the opioid epidemic?* December 7, 2018 February 13, 2019 [cited 2019 December 12]; Available from: https://www.brookings.edu/blog/up-front/2018/12/07/research-roundup-what-does-the-evidence-say-about-how-to-fight-the-opioid-epidemic/.
6. Wen, H., et al., *Prescription Drug Monitoring Program Mandates: Impact On Opioid Prescribing And Related Hospital Use.* Health Affairs, 2019. **38**(9): p. 1550-1556.
7. Johnson, H., et al., *Decline in Drug Overdose Deaths After State Policy Changes - Florida, 2010-2012.* CDC MMWR Weekly, July 4, 2014. **63**(26): p. 569-574.
8. National Academies of Sciences, E., and Medicine; Health and Medicine Division; Board on Health Sciences Policy; Committee on Pain Management and Regulatory Strategies to Address Prescription Opioid Abuse. *Pain Management and the Opioid Epidemic: Balancing Societal and Individual Benefits and Risks of Prescription Opioid Use*. July 13, 2017; Available from: https://www.ncbi.nlm.nih.gov/books/NBK458660/.
9. Fink, D.S., et al., *Association Between Prescription Drug Monitoring Programs and Nonfatal and Fatal Drug Overdoses: A Systematic Review.* Annals of Internal Medicine, 2018. **168**(11): p. 783-790.
10. Doctor, J.N., et al., *Opioid prescribing decreases after learning of a patient's fatal overdose.* Science, 2018. **361**(6402): p. 588-590.
11. Mattick, R.P., et al., *Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence.* Cochrane Database of Systematic Reviews, 2014(2).
12. Sordo, L., et al., *Mortality risk during and after opioid substitution treatment: systematic review and meta-analysis of cohort studies.* BMJ, 2017. **357**: p. j1550.
13. Carroll, J., T. Green, and R. Noonan. *Evidence-Based Strategies for Preventing Opioid Overdose: What's Working int he United States*. 2018 [cited 2019 December 12]; Available from: https://www.cdc.gov/drugoverdose/pdf/pubs/2018-evidence-based-strategies.pdf.
14. D'Onofrio, G., et al., *Emergency Department–Initiated Buprenorphine/Naloxone Treatment for Opioid Dependence: A Randomized Clinical Trial.* JAMA, 2015. **313**(16): p. 1636-1644.
15. Rich, J.D., et al., *Methadone continuation versus forced withdrawal on incarceration in a combined US prison and jail: a randomised, open-label trial.* The Lancet, 2015. **386**(9991): p. 350-359.

16. Wen, H., J.M. Hockenberry, and H.A. Pollack, *Association of Buprenorphine-Waivered Physician Supply With Buprenorphine Treatment Use and Prescription Opioid Use in Medicaid Enrollees.* JAMA Network Open, 2018. **1**(5): p. e182943-e182943.

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.

The goal of this measure is to accurately quantify emergency department use due to opioid overdose among Medicare beneficiaries. Accurate and timely measurement of opioid overdose can serve several purposes: it can provide stakeholders with a tool to measure the burden of opioid overdose within a community, it enables comparison among geographies, and it allows for tracking trends and improvement over time within entities. Measure results could be used to focus resources on communities most in need and to encourage investment in reducing opioid overdose, in addition to incentivizing innovation and systems improvement.

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 used Medicare Claims data from 2017 from 25 states for measure development and our main testing (reliability, validation). We additionally evaluated performance in 2 additional years, 2016 and 2018. Please see the testing form for a full description of the data source.

Below are state-level results for each year, in ED visits per 1000 person-years (mean (SD); Range [min, max]):

2017: 1.11 (0.36); Range 0.53-1.80

2018: 0.92 (0.33); Range 0.42-1.71

Below are county-level results for each year (mean (SD); Range[Min-max]):

2017: 1.55 (1.11); Range 0.59-6.22

2018: 1.44 (1.23); Range 0.48-6.44

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.

Over the past decade, the United States has seen marked increase in opioid use, opioid overdose, and death due to opioid overdose. Between 1999 and 2017, drug overdose deaths in the US, most of which were due to opioids, rose approximately four-fold, from 16,849 to 70,237 (1). Concurrent with this rise in opioid-related deaths, emergency department visits for opioid overdose have also risen sharply. Hasegawa et al. used nationally representative data to quantify ED visits for opioid overdose and estimated that the rate of ED visits for opioid overdose rose by nearly four-fold between 1993 and 2010 (2). A report by the Agency for Healthcare Research and Quality (AHRQ) noted a doubling in the national rate of opioid-related ED visits between 2005 and 2014, and 64% increase in the rate of hospital admissions (3). This sharp rise in opioid use and overdose

has generated national concern, and in 2017, the opioid crisis was declared a federal Public Health Emergency (4).

Although the nation as a whole has seen a marked rise in opioid overdose, rates of opioid-related ED visits vary considerably by state. For example, in 2014, Massachusetts had a rate of 450 visits per 100,000 population while Iowa had a rate of 45 per 100,000 (3). This variation suggests that a variety of regional level factors may be important, and such factors could be the targets of health care and public health interventions.

Current measures of opioid overdose, which have been used to report these performance gaps, have been developed for a variety of research and public health purposes. Our goal was to build on this work to develop a measure of opioid overdose that had improved face validity, balanced specificity with inclusivity, and was tested for validity and reliability at multiple geographic levels. County level measurement is particularly useful as it would allow states to target resources to counties hardest hit and to track improvements over time with greater geographic specificity and granularity.

References

1. https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates, accessed 2/7/19

2. Hasegawa K, Espinola JA, Brown DF, Camargo CA, Jr. Trends in U.S. emergency department visits for opioid overdose, 1993-2010. Pain Med. 2014;15(10):1765-1770.

3. Weiss AJ, Elixhauser A, Barrett ML, Steiner CA, Bailey MK, O´Malley L. Opioid-Related Inpatient Stays and Emergency Department Visits by State, 2009-2014. 2017. https://www.hcup-us.ahrq.gov/reports/statbriefs/sb219-Opioid-Hospital-Stays-ED-Visits-by-State.jsp. Accessed September 11, 2017.

4 https://www.hhs.gov/about/news/2017/10/26/hhs-acting-secretary-declares-public-health-emergency-address-national-opioid-crisis.html, accessed 2/7/19

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.*

N/A

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

Multiple studies have described demographic risk factors for overdose, including age, sex, and race(1). In addition, although opioid use disorder and opioid overdose are seen in all strata of society, opioid use disorder is more common among people living in poverty (2). Further, treatment is not equally accessible to all. Treatment for opioid use disorder is less available in some states and regions because of restrictive legislation and/or provider availability, and some settings such as within the criminal justice system (2, 3).

References

1. Brady JE, Giglio R, Keyes KM, DiMaggio C, Li G. Risk markers for fatal and non-fatal prescription drug overdose: a meta-analysis. Inj Epidemiol. 2017;4(1):24.

2. Dasgupta N, Beletsky L, Ciccarone D. Opioid Crisis: No Easy Fix to its Social and Economic Determinants. Am J Public Health. 2018 Feb; 108(2) 182-186.

3. Abraham AJ, Andrews CM, Yingling ME, Shannon J. Geographic Disparities in Availability of Opioid Use Disorder Treatment for Medicaid Enrollees.

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.)*

N/A

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

This is not an eMeasure **Attachment:** Del18hHOP5MarylandOpiodDataDictionary01042019.xlsx,Del18hHOP5MarylandOpiodTestingForm01042019-636824484802548809.docx,Del18hHOP5MarylandOpiodTestingForm01042019.docx,Del18hHOP5MarylandOpiodTestingForm01222019.docx

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

Attachment **Attachment:** 7.31.19_Data_Dictionary_v1.0.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.

No

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.

N/A

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).

The numerator is comprised of incident outcome events, defined as opioid overdoses that result in emergency department use, within the population residing in a specific geography.

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 numerator is comprised of outcome events, i.e., emergency department (ED) visits for opioid overdose. This numerator includes all overdose events that result in treatment in the emergency department in the measured population within a one-year measurement period. The measured population is defined below in Section 5.6 and 5.7.

To capture overdose events, the measure first identifies all ED visits for the measured population using a validated algorithm (Venkatesh, 2017). Details of this algorithm are included in the measure Data Dictionary. From among these ED visits, the measure then identifies visits for opioid overdose using a set of ICD-10 diagnostic codes. Opioid overdose is defined by the presence of a diagnostic code indicating opioid poisoning such as T400X4A (Poisoning by opium, undetermined, initial encounter). This code can appear as either a principal discharge diagnosis or a secondary diagnosis. The measure outcome definition excludes ICD-10 codes indicating intentional overdose or assault. Only diagnostic codes indicating an initial encounter are included. See the Data Dictionary for the full set of codes comprising the outcome definition.

Opioid overdoses resulting in an ED visit are included regardless of final disposition (e.g., admission, discharge etc.) or vital status (i.e., alive or deceased) at discharge. Repeat events for individual patients are also included, as the goal of the measure is to capture all unintentional opioid overdoses in the measured population. Indeed, an overdose is a risk factor for subsequent overdose, and has been proposed as an important opportunity for intervention (Larochelle, 2018). Thus, including repeat events is important for measuring opioid overdose as a population health measure. Outcome events are attributed to a geography based on a person's residence, not based on the emergency department in which an individual seeks care.

Reference

(1) Venkatesh AK, Mei H, Kocher KE, et al. Identification of Emergency Department Visits in Medicare Administrative Claims: Approaches and Implications. Academic Emergency Medicine. 2017;24(4):422-431.

(2) Larochelle MR, Bernson D, Land T, Stopka TJ, Wang N, Xuan Z, et al. Medication for Opioid Use Disorder After Nonfatal Opioid Overdose and Association With Mortality: A Cohort Study. Ann Intern Med. [Epub ahead of print 19 June 2018]169:137–145. doi: 10.7326/M17-3107

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

The denominator consists of all enrolled Medicare Fee-For-Service (FFS) beneficiaries with Parts A or B, aged 18 and older residing in a measured geography (either a county or a state) during a one-year period.

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 denominator includes all Medicare beneficiaries enrolled in Medicare Part A or B who are at least 18 years of age residing in the measured geography.

The denominator reflects the size of the population in which overdose events occur, measured in person-years. Person-years is calculated by summing the fraction of a year each eligible beneficiary is enrolled in

Medicare over the entire measured population. For example, one person enrolled for a year would contribute one person-year to the denominator. One person enrolled for 6 months would contribute 0.5 person-years. These enrollment periods are summed over the entire eligible population to calculate the total person-years for. Periods during which beneficiaries are not enrolled are considered periods during which the outcome cannot be measured and therefore are not included in the denominator.

The measure is designed to be used as a population health measure and has been tested at two different geographic levels, the county and the state. Eligible beneficiaries are assigned to geographies based on place of residence. Thus, individuals contribute to the denominator and the numerator based on residence rather than where the event took place.

Identifying emergency department visits requires information from both inpatient and outpatient claims which are covered by Medicare Parts A and B respectively. In order to be maximally inclusive, the measure includes all beneficiaries with either Part A or B, rather than requiring that beneficiaries have Parts A and B. Limiting the measure to beneficiaries who have Parts A and B would exclude individuals with observable outcome events. For example, beneficiaries with Part A would have observable outcome events if they are admitted to the hospital for an opioid overdose while those with Part B would have an observable outcome event if they were seen only in the emergency department. Although this approach may miss some outcome events for beneficiaries with only Parts A or B, it allows the measure to be maximally inclusive of both the measured population and potential outcome events.

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

None

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.)*

None

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.)*

None

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

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

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

Better quality = Lower score

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

This measure estimates the rate of emergency department visits for opioid overdose events.

Events are measured per 1,000 person-years among Medicare beneficiaries 18 years of age or older residing in the geography being measured. The calculation is detailed below:

1. Identify target population: (Medicare Part A or B enrollment, age 18 years or older residing in a measured geography in the measured timeframe)

2. Calculate enrollment period for each eligible beneficiary

3. Calculate total person-years for the geography of interest by summing person-years among included beneficiaries

4. Calculate numerator (overdose events resulting in an emergency department visit according to the measure definition)

5. Calculate ratio of numerator to denominator * 1,000

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

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

N/A

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

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

N/A

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

If other, please describe in S.18.

Claims

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

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

Data for measure development and testing were collected from the Medicare Fee-For-Service (FFS) claims data. We used a 100% sample of Medicare beneficiaries for 25 states and used both inpatient and outpatient claims to identify emergency department visits. We used data from 2017 for measure development and validation.

Medicare claims: This data source contains claims data for FFS inpatient and outpatient services including Medicare inpatient hospital care, outpatient hospital services (which includes emergency services) and physician services (carrier claims).

Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on Medicare enrollment status during the measurement period.

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)*

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

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.)*

N/A

2. Validity – See attached Measure Testing Submission Form

7.31.19_nqf_testing_attachment_7.1_v1.0.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.

No

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.

No

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Acute Care Use Due to Opioid Overdose
Date of Submission: 8/1/2019
Type of Measure:

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

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: Click here to describe	☐ other: Click here to describe

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*).

To develop and test the measure, we used a dataset of Medicare claims data. This dataset includes a 100% sample of Medicare Inpatient, Outpatient, and Carrier (Part B Physician) claims from 25 states. We also used the Beneficiary Enrollment Database to establish the time during which each beneficiary was enrolled in Medicare.

In addition, for measure validation, we used published, aggregated data from the Nationwide Emergency Department Sample (NEDS) and National Inpatient Sample (NIS) and CDC data on opioid overdose deaths, compiled by the Kaiser Family Foundation.

1.3. What are the dates of the data used in testing? 2017, with additional meaningful differences testing performed in 2018.

2017, 2018

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

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

1.5. How many and which <u>measured entities</u> 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*)

The measure has been tested at two geographic levels: the state and county. Because this measure has primarily been developed for use in the state of Maryland, we tested it in Maryland and in twenty-four other states: Arizona, California, Florida, Georgia, Iowa, Illinois, Indiana, Kansas, Kentucky, Maine, Michigan, Minnesota, Missouri, Montana, North Carolina, North Dakota, Nebraska, Nevada, Oregon, South Dakota, Tennessee, Texas, Wisconsin, and Wyoming. We chose these states because they are included in the 2017 National Emergency Department Sample (NEDS) and the National Inpatient Sample (NIS) which was used for validation.

We also tested the measure for use at the county level. Specifically, we tested it in all 24 counties and Baltimore city (considered a county equivalent) in the state of Maryland.

1.6. How many and which <u>patients</u> 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*)

Because this is a population health measure, it includes all Medicare beneficiaries aged 18 and older enrolled in Medicare Part A or B residing in a measured geographic area during a one-year measurement period. The measure denominator, which reflects the size of the measured population, is expressed in person-years, rather than as a number of individuals. Person-years is calculated by summing the time each eligible beneficiary is enrolled over the entire measured population. For example, one person enrolled for a year would contribute one person-year to the denominator. One person enrolled for 6 months would contribute 0.5 person-years. These enrollment periods are summed over the entire eligible population to calculate the total person-years for the measured population. Periods during which beneficiaries are not enrolled are periods during which the outcome cannot be measured and therefore are not included in the denominator.

Because our denominator is a population expressed in person-years, it is more informative to report the number of person-years included in the measure, which is less than or equal to the number of individuals included in the denominator in a one-year period. See Section 1.7 for details.

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.

Dataset	Applicable Section in the Testing Attachment	Description of Dataset
Medicare Claims	Section 2a2 Reliability Testing Section 2b1 Data Element & Measure Score Validity 2b4 Meaningful Differences	Dates of Data: 1/1/2017 – 12/31/2017 Number of States: 25 Range: 96,927 - 3,429,365 Mean: 886,412 Person-years by State (2017): Maryland: 863,860 Arizona: 765,550 California: 3,429,365 Florida: 2,494,132 Georgia: 1,052,190 Iowa: 486,660 Illinois: 1,609,944 Indiana: 879,436 Kansas: 426,934 Kentucky: 632,808 Maine: 230,904 Michigan: 1,250,663 Minnesota: 411,865 Missouri: 801,344 Montana: 170,295 North Carolina: 1,278,766 North Dakota: 100,547 Nebraska: 284,299 Nevada: 322,098 Oregon: 441,974 South Dakota: 129,739 Tennessee: 821,243 Texas: 2,513,350 Wisconsin: 665,415 Wyoming: 96,927 Number of Counties in Maryland: 25 Range: 6,050 - 136,139 Mean: 35,994 Number of Person-years by county in Maryland (2017): Allegany: 16,030 Anne Arundel: 77,692 Baltimore: 128,248 Baltimore City: 87,221 Calvert: 14,106 Caroline: 6,100 Carroll: 28,762 Cecil: 17,003

Dataset	Applicable Section in the Testing Attachment	Description of Dataset
		Charles: 19,411 Dorchester: 7,075 Frederick: 37,637 Garrett: 5,891 Harford: 40,619 Howard: 38,890 Kent: 6,050 Montgomery: 136,139 Prince George's: 101,273 Queen Anne's: 8,110 Saint Mary's: 15,184 Somerset: 4,260 Talbot: 10,188 Washington: 26,431 Wicomico: 17,546 Worcester: 13,994 Demographics of individuals included in the Medicare dataset: Mean age 70.5 (SD 12.8) Sex: 54% female, 46% male Race/ethnicity: 80% White, 10% Black, 2.7 % Asian, 3.5% Hispanic, 1.9% other, 0.4% Native American
Nationwide ED Sample/National Inpatient Sample	Section 2b1 Data Element & Measure Score Validity	Dates of Data: 2017 Number of States: 24 Emergency department visits that result in discharge form the emergency department are included in the NEDS. Visits that result in hospital admission are captured in the NIS. NEDS and NIS data are meant to be a comprehensive database of all ED visits rather than a sample. Accordingly, the denominator used to derive rates in the NEDS is the entire adult population residing within a state.
Kaiser Family Foundation	Section 2b1 Data Element & Measure Score Validity	Dates of Data: 2017 Number of States: 25 The Kaiser Family Foundation has used data from The National Vital Statistics Center/CDC Wonder to produce age-adjusted death rates by state each year. Deaths are classified according to ICD-10 multiple cause of death codes and these codes are used to establish whether opioid overdose was the cause of death. The denominator is the state population. Rates are age adjusted by applying age-

Dataset	Applicable Section in the Testing Attachment	Description of Dataset
		specific death rates to the 2000 U.S. standard population age distribution.

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.

We analyzed the following community characteristics at the state and county level: Race (percent white) ethnicity (percent non-Hispanic), median household income, percent of residents below the poverty line, and percent of residents living in a rural area (at the state level only). See Section 2b3.2. for a discussion of testing results.

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*)

Measure Score Reliability

We estimated the reliability at the level of the measured geographic area. For reliability testing, we believe it is meaningful to consider the notion of "unit" reliability, that is, the reliability with which individual units (here, states or counties) are measured. This is because the reliability of any one geography's measure score will vary depending on the number of overdose events. Geographies with more events will tend to have more reliable scores, while geographies with fewer will tend to have less reliable scores. Therefore, we use the formula presented by Adams and colleagues (2010).

We also performed split sample reliability testing. For spilt sample reliability testing, we randomly divided the sample into two parts and compared measure outcomes in each half of the data using a correlation coefficient.

Additional Information

In constructing the measure in Medicare Fee-For-Service (FFS) data, we aim to utilize only those data elements from claims data that have both face validity and reliability. We avoid the use of fields that are thought to be coded inconsistently across regions. Specifically, we used fields that are consequential for payment and which are audited. We identify such variables through empiric analyses and our understanding of the CMS auditing and billing policies, and we seek to avoid variables which do not meet this standard.

In addition, CMS has in place several auditing programs used to assess overall claims coding accuracy, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in our

measures, including diagnosis and procedure codes and other elements that are consequential for payment.

Reference

1. Adams J, Mehrota, A, Thoman J, McGlynn, E. (2010). Physician cost profiling – reliability and risk of misclassification. NEJM, 362(11): 1014-1021.

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*)

Measure Score Reliability:
Among the 25 states evaluated (MD, AZ, CA, FL, GA, IA, IL, IN, KS, KY, ME, MI, MN, MO, MT, NC, ND, NE, NV, OR, SD, TN, TX, WI, WY), Adams reliability ranged from 0.92-0.99 with a mean of 0.98. Among counties in Maryland, Adams reliability ranged from 0.60-0.99 with a mean of 0.89. Among counties, only a single county had a reliability score below 0.7.

Split sample reliability testing indicated a high correlation (r=0.94) between split samples at the state level. At the county level, measure results between the split samples had a correlation coefficient of 0.87.

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?*)

Testing indicated that, at both geographic levels (county and state), reliability was nearly universally ≥0.7. A single Maryland county had a reliability score of 0.60, with all others above 0.7, indicating acceptable reliability. All states had a reliability score >0.92 indicating high reliability.

Split sample reliability testing indicated a strong correlation between samples at both the county and state levels.

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)*

Measure Score: Construct Validity
The outcome definition used in this measure uses diagnostic codes to identify emergency department visits for opioid overdose. Specifically, the measure requires a principal or secondary diagnosis code that explicitly indicates *opioid poisoning* associated with an emergency department visit (see data dictionary for code list). The outcome definition includes only diagnosis codes indicating unintentional or undetermined poisonings and excludes poisoning related to assault or self-harm. The measure outcome definition also only includes diagnostic codes indicating an initial encounter.

Several studies have evaluated the sensitivity, specificity and positive predictive value of these opioid poisoning codes, using chart review with a prespecified case definition as a gold standard. In general, this group of codes is highly specific for opioid overdose (99%), but has a sensitivity of 25% (Rowe, 2017). The positive predictive value of these codes for opioid overdose in a very large general medical population was 81% (Green, 2017). A second, smaller study reported a positive predictive value of 70% in an emergency medicine population (Reardon, 2016). Although these studies primarily used ICD-9 codes, ICD-9 codes map

directly to ICD-10 codes with very similar descriptors. Thus, the published literature helps to support the validity of this group of codes for identifying opioid overdose with high specificity and acceptable positive predictive value.

Measure Score Validity - Empirical Testing Through Comparison to Independent Data Sources:
There is no established gold standard for measuring opioid overdose from claims data. Therefore, we compared the proposed measure to two other sources: a claims-based measure of opioid-related acute care use independently developed and reported by the Agency for Health Care Research and Quality (AHRQ) and to state-level opioid overdose death rates from the Centers for Disease Control and Prevention. These comparisons are described in detail below.

NEDS/NIS Comparison:
We compared state-level results from this measure to the rate of opioid-related acute care use reported by AHRQ using data from the Nationwide Emergency Department Sample (NEDS) and the National Inpatient Sample (NIS) (HCUP Fast Stats, 2018). The NEDS/NIS is a database of emergency department (ED) visits and inpatient admissions maintained by the Agency for Healthcare Research and Quality. Emergency department visits that result in discharge from the ED are captured in the NEDS while visits resulting in admission are captured in the NIS. Together, the databases offer a comprehensive record of acute care visits for a state. NEDS and NIS use administrative data and include patients of all payer types and diagnoses are based on ICD-10 codes associated with ED visits or hospitalizations.

The Agency for Healthcare Research and Quality (AHRQ) has developed a definition of opioid-related acute care use that overlaps considerably with the definition used in this measure (Weiss, 2016). However, the AHRQ definition is broader and captures visits for opioid-associated symptoms, withdrawal, or related to opioid use disorder. AHRQ has used this definition to estimate the number of opioid-related acute care visits in the NEDS/NIS. In this measure, visits are reported as a rate, with opioid-related ED visits as the numerator and the state population as the denominator. Because the two databases are mutually exclusive, the sum of ED visits and inpatient hospitalizations approximates the total number of opioid-related ED visits per population over a one-year period, assuming the great majority of people admitted with an opioid overdose are initially seen in the emergency room. We compared the rate of ED visits given by this measure to the rate in the proposed measure.

Opioid Overdose Death Rates
In addition, we also compared the rate of opioid overdose given by the proposed measure to opioid overdose death rates from the Centers for Disease Control/Kaiser Family Foundation (The Henry J. Kaiser Family Foundation, 2017). State overdose death rates are based on data from the National Vital Statistics Program and are age-adjusted to the US population overall. Cause of death is based on ICD-10 codes.

In comparing the proposed measure to overdose deaths, we acknowledge that overdose death and overdose may not completely track together. Indeed, communities with better access to resources such as naloxone and EMS services may have higher overdose rates and lower mortality because a larger proportion of those who overdose survive. However, in general, ED visits for opioid overdose and overdose deaths likely originate from the same underlying epidemic and are likely to track together. Accordingly, we compared age-adjusted opioid overdose death rates at the state level in 2017 for the 25 states in our sample to measure outcome rates in 2017.

Measure Face Validity:
To assess face validity, we convened a meeting of clinicians with expertise in opioid use and emergency medicine. This group included four practicing emergency medicine physicians, all with research expertise in opioid use disorder and in quality measurement, and two general internists with expertise in opioid use disorder, treatment, and policy. Members of this group are considered national and international experts in opioid use disorder, measurement, treatment, and policy. Several group members have served on state or

national policy advisory committees and head large, multicenter collaborations to track and address opioid use disorder. All members of this group were faculty members at the Yale School of Medicine.

We presented information on the measure context, potential uses, and measure specifications to this expert panel. We also presented a version of this measure with a much broader outcome definition that may be more sensitive but less specific. This broader definition includes all ICD-10 codes included in the present measure. It also captures patients who present to the emergency department with diagnoses suggesting opioid use *and* symptoms consistent with overdose (for example F11.20 "Opioid dependence, uncomplicated" with J96.01 "Acute respiratory failure with hypoxia"). Panel members were given the opportunity for discussion and to ask questions.

Members of our clinical expert panel indicated that that the proposed measure was highly specific to opioid overdose. Given the option of a broader versus narrower measure, the panel felt that maintaining measure specificity was preferable to a broader, more sensitive but less specific measure. However, the group recommended using the broader measure for internal tracking purposes, as a way to detect potential measure gaming or changes that are related to coding practices and not actual changes in overdose rates. At the conclusion of the panel, 5 members voted on the face validity of the measure. All 5 voting members ranked the measure a 4 on a Likert scale, with 1 indicating members strongly disagreed and 5 indicating members strongly agreed that the measure had adequate face validity.

References

1. Green, CA, Perrin, NA, Janoff, SL, Campbell, CI, Chilcoat, HD, Coplan, PM. Assessing the accuracy of opioid overdose and poisoning codes in diagnostic information from electronic health records, claims data, and death records. Pharmacoepidemiol Drug Saf. 2017; 26(5): 509-517.

2. HCUP Fast Stats. Healthcare Cost and Utilization Project (HCUP). October 2018. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/faststats/opioid/opioiduse.jsp?radio3=on&location1=MD&characteristic1=06&setting1=ED&location2=&characteristic2=01&setting2=IP&expansionInfoState=hide&dataTablesState=show&definitionsState=show&exportState=hide.

3. Prescription Opioid Overdose Deaths and Death Rate per 100,000 Population (Age-Adjusted). The Henry J. Kaiser Family Foundation; 2017. https://www.kff.org/other/state-indicator/opioid-overdose-death-rates/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D. Accessed July 10, 2019.

4. Reardon JM , Harmon KJ , Schult GC , et al . Use of diagnosis codes for detection of clinically significant opioid poisoning in the emergency department: A retrospective analysis of a surveillance case definition. BMC Emerg Med 2016;16:11.doi:10.1186/s12873-016-0075-4

5. Rowe, C, Vittinghoff, E, Santos, GM, Behar, E, Turner, C, Coffin, PO. Performance measures of diagnostic codes for detecting opioid overdose in the emergency department. Academic emergency medicine: Official journal of the society for. Acad Emerg Med. 2017; 24(4): 475-483.

6. Weiss AJ, Elixhauser A, Barrett ML, Steiner CA, Bailey MK, O'Malley L. Opioid-Related Inpatient Stays and Emergency Department Visits by State, 2009-2014. (2016). HCUP Statistical Brief #219. Agency for Healthcare Research and Quality, Rockville, MD. http://www.hcup-us.ahrq.gov/reports/statbriefs/sb219-Opioid-Hospital-Stays-ED-Visits-by-State.pdf.

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

In 2017, the rate of ED visits for opioid overdose in the proposed measure ranged from 0.86-1.84 per 1,000 person-years among the 25 states tested. In NEDS/NIS data, which is an all-payer population, rates ranged from 2.27-9.60 per 1,000 population per year. Correlation between the two measures was high (r=0.74).

Correlation between the proposed measure outcome rate and opioid overdose death rate was also high (0.74) at the state level among 25 states tested.

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?*)

Review of the published literature suggests that the codes that comprise this measure are highly specific for opioid overdose and even in an average-risk population, have a high positive predictive value.

Our empirical validity testing results demonstrate that the proposed measure, which is designed to capture ED visits for opioid overdose, is correlated with two independent measures: a broader measure of opioid-related acute care use, and deaths due to opioid overdose. These analyses suggest that our measure tracks with other indicators of population-level opioid-related morbidity and mortality.

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*)

N/A

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*)

N/A

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*)

N/A

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** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

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.

N/A

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.

This measure is not risk adjusted for two conceptual reasons. First, although the risk of opioid overdose varies according to patient demographic characteristics (Kaiser Family Foundation, 2018), this variation reflects social rather than biological differences. Thus, risk adjusting may obscure important health disparities. Second, because differences in opioid overdose rates are due to social rather than biological differences, risk adjusting would tacitly allow for higher rates in some populations and would perpetuate inequities in the health of communities.

Although we believe that opioid overdose should be fully preventable in all populations, we acknowledge that opioid use disorder and opioid overdose are produced by the complex interplay of biological and social phenomena, and there are challenges and complexities when comparing communities with very different histories and resources. To better understand this, we evaluated the relationship between area-level sociodemographic characteristics and measure outcome rates. Among the twenty five states in our testing sample, we observed low or moderate correlations between the rate of ED visits for opioid overdose and percent of the population that is white (r=-0.40), percent non-Hispanic (r=0.09), median household income (r=0.05), percent of the population living in poverty (r=0.25) and percent of the population living in rural areas (r=-0.22). Among counties in Maryland, ED use for opioid overdose was negatively correlated with median household income (r=-0.64) and positively correlated with percent of the population living in poverty (r=0.80).

These results suggest that on the state level, the relationship between state-level wealth and opioid overdose is not strong. However, within a state, at the county level, there is a stronger relationship between area-level wealth and population-based rates of ED visits for opioid overdose. Although areas with fewer economic resources may have higher rates of opioid overdose, we believe that adjusting for area socioeconomic factors would obscure and perpetuate disparities. Tracking entities over time, however, would allow communities to improve while acknowledging that history, culture, and resources vary and play a complex role in the opioid epidemic. Comparing entities to one another would identify areas hardest hit by the opioid epidemic and most in need of resources to improve overdose rates.

Reference

1. (2016). "Opioid Overdose Deaths by Race/Ethnicity." State Health Facts. Retrieved December 11, 2018, from https://www.kff.org/other/state-indicator/opioid-overdose-deaths-by-raceethnicity/?dataView=2&currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D#notes

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?

N/A

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)**
This decision was made on conceptual grounds, as described in section 2b3.2

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

N/A

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.**

N/A

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*)

N/A

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*)**:**

N/A

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

N/A

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

N/A

2b3.9. Results of Risk Stratification Analysis:

N/A

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*)

N/A

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*)

N/A

_____

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)*

Consistent with other quality measures, we calculate 95% confidence interval estimates for the opioid overdose rate to characterize the amount of uncertainty associated with the outcome rate, compare the interval estimate to the state or county average rate, and categorize states or counties as "higher than," "less than," or "no different than" the average (Sahai, 1996). To compare averages, we used a one-sample t-test.

In addition to comparing entities to each other within a year, we also compared entities to themselves from one year to the next. Here, we used a generalized linear model with a Poisson distribution and a population offset to evaluate whether incidence of opioid overdose in a measured population has changed overtime. Similarly, we categorized differences between states or counties over time as "higher than", "less than" or "no different than" prior.

Reference

1. Sahai H, Khurshid A (1996) *Statistics in epidemiology: methods, techniques, and applications*. Boca Raton, FL: CRC Press, Inc.

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*)

Meaningful difference was defined as values statistically above or below the average for the measured group when comparing groups. When comparing changes within an entity, we defined a meaningful difference as a difference between measures that is statistically significantly different from zero.

State measurements (25 states in 2017):
12 states below average
3 states same as average
10 states above average

State Rate per 1,000 person-years [Confidence Interval] (2017):
- Maryland: 1.801 [1.714 - 1.893]
- Arizona: 0.863 [0.800 - 0.932]
- California: 1.013 [0.980 - 1.047]
- Florida: 1.253 [1.210 - 1.298]
- Georgia: 1.164 [1.101 - 1.231]
- Iowa: 0.567 [0.504 - 0.638]
- Illinois: 0.995 [0.948 - 1.045]
- Indiana: 1.379 [1.304 - 1.459]
- Kansas: 1.024 [0.932 - 1.124]
- Kentucky: 1.596 [1.501 - 1.698]
- Maine: 1.087 [0.961 - 1.230]
- Michigan: 1.841 [1.768 - 1.918]
- Minnesota: 1.377 [1.268 - 1.495]
- Missouri: 1.300 [1.224 - 1.382]
- Montana: 0.810 [0.686 - 0.957]
- North Carolina: 1.270 [1.210 - 1.333]
- North Dakota: 0.597 [0.463 - 0.768]
- Nebraska: 0.598 [0.515 - 0.695]
- Nevada: 1.388 [1.265 - 1.522]
- Oregon: 0.996 [0.907 - 1.093]
- South Dakota: 0.532 [0.420 - 0.673]
- Tennessee: 1.354 [1.277 - 1.436]
- Texas: 0.874 [0.838 - 0.911]
- Wisconsin: 0.981 [0.909 - 1.060]
- Wyoming: 1.011 [0.830 - 1.232]

County measurements (25 counties in 2017):
2 (8%) above average and 9 below average (36%)

County Rate per 1,000 person-years [Confidence Interval] (2017):
- Allegany: 1.497 [1.004 - 2.233]
- Anne Arundel: 1.364 [1.128 - 1.650]
- Baltimore: 2.448 [2.192 - 2.734]
- Baltimore City: 6.145 [5.648 - 6.687]
- Calvert: 0.851 [0.483 - 1.498]
- Caroline: 1.639 [0.882 - 3.045]
- Carroll: 1.773 [1.348 - 2.333]
- Cecil: 1.882 [1.331 - 2.660]
- Charles: 1.391 [0.954 - 2.028]
- Dorchester: 1.838 [1.068 - 3.163]
- Frederick: 0.638 [0.427 - 0.951]
- Garrett: 1.018 [0.458 - 2.266]
- Harford: 1.329 [1.018 - 1.736]
- Howard: 0.771 [0.539 - 1.103]
- Kent: 0.826 [0.344 - 1.985]
- Montgomery: 0.485 [0.381 - 0.617]
- Prince George's: 0.987 [0.812 - 1.201]
- Queen Anne's: 1.110 [0.578 - 2.132]
- Saint Mary's: 1.515 [1.007 - 2.279]

Somerset: 2.113 [1.100 - 4.058]
Talbot: 0.589 [0.265 - 1.311]
Washington: 1.816 [1.369 - 2.409]
Wicomico: 2.166 [1.576 - 2.975]
Worcester: 0.929 [0.540 - 1.600]

Meaningful Differences between 2017 and 2018 (State-level Data)

State	2017 Rate	2018 Rate	p value
Maryland	1.80	1.71	0.16
Arizona	0.863	0.731	0.0035
California	1.013	0.917	<.0001
Florida	1.253	1.088	<.0001
Georgia	1.164	1.037	0.0057
Iowa	0.567	0.418	0.001
Illinois	0.995	0.994	0.97
Indiana	1.379	0.995	<.0001
Kansas	1.024	0.772	0.001
Kentucky	1.596	1.186	<.0001
Maine	1.087	1.069	0.85
Michigan	1.841	1.617	<.0001
Minnesota	1.377	1.007	<.0001
Missouri	1.300	0.977	<.0001
Montana	0.810	0.552	0.0035
North Carolina	1.270	0.976	<.0001
North Dakota	0.597	0.425	0.09
Nebraska	0.598	0.444	0.0111
Nevada	1.388	1.112	0.0017
Oregon	0.996	0.864	0.0408
South Dakota	0.532	0.476	0.52
Tennessee	1.354	1.153	0.0003
Texas	0.874	0.744	<.0001
Wisconsin	0.981	0.791	0.0002
Wyoming	1.011	0.822	0.17

Meaningful Differences between 2017 and 2018 (County-level Data)

County	2017 Rate	2018 Rate	p value
Allegany	1.497	2.039	0.25
Anne Arundel	1.364	1.347	0.92
Baltimore	2.448	2.084	0.05
Baltimore City	6.145	6.436	0.44
Calvert	0.851	1.502	0.11
Caroline	1.639	0.813	0.2
Carroll	1.773	1.504	0.42
Cecil	1.882	2.549	0.19
Charles	1.391	1.386	0.99
Dorchester	1.838	0.562	0.0382
Frederick	0.638	0.649	0.95
Garrett	1.018	1.006	0.98
Harford	1.329	0.993	0.16
Howard	0.771	1.137	0.1

Kent	0.826	0.494	0.48
Montgomery	0.485	0.484	0.99
Prince George's	0.987	0.671	0.0133
Queen Anne's	1.110	0.842	0.58
Saint Mary's	1.515	1.308	0.63
Somerset	2.113	1.866	0.8
Talbot	0.589	0.970	0.33
Washington	1.816	2.494	0.09
Wicomico	2.166	0.616	0.0002
Worcester	0.929	0.777	0.66

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 variation in rates suggests that there are meaningful differences across counties and states in the rates of opioid overdose resulting in emergency department use. Our results also suggest that the measure can be used to identify meaningful differences with states and counties over time. As expected, there are fewer statistically significant differences in counties in a one-year period than among states. However, we did observe some statistically significant differences, even in a one-year interval.

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*)

N/A

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*)

N/A

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*)

N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

N/A

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*)

N/A

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

N/A

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

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

N/A

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

Attachment:

3c. Data Collection Strategy

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

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

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

Administrative data are routinely collected as part of the billing process. Because completion of claims is required for hospital reimbursement, there is little missing data. The measures do not require any additional data collection.

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)*.

There are no fees associated with the use of this measure

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Health/Disease Surveillance	

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

This measure is not in current use.

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

N/A. This measure is not publicly reported.

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.*)

This measure would be used to track Maryland's progress on reducing opioid overdoses under the Maryland Total Cost of Care Model.

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.

N/A

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.

N/A

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.

N/A

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

N/A

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

N/A

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.

N/A

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.

N/A

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.

N/A

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

N/A

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?
5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

N/A

5b. Competing Measures

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

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

N/A

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested

information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment:

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare & Medicaid Services (CMS)

Co.2 Point of Contact: Lein, Han, lein.han@cms.hhs.gov, 410-786-0205-

Co.3 Measure Developer if different from Measure Steward: Yale New Haven Health Services Corporation/Center for Outcomes Research and Evaluation (YNHHSC/CORE)

Co.4 Point of Contact: Darinka, Djordjevic, darinka.djordjevic@yale.edu, 203-764-5700-

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.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

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

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

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: