NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 3389

Corresponding Measures:

De.2. Measure Title: Concurrent Use of Opioids and Benzodiazepines (COB)

Co.1.1. Measure Steward: PQA, Inc.

De.3. Brief Description of Measure: The percentage of individuals >=18 years of age with concurrent use of prescription opioids and benzodiazepines during the measurement year.

A lower rate indicates better performance.

1b.1. Developer Rationale: Overdose deaths involving prescription opioids were five times higher in 2016 than in 1999, and more than 200,000 people have died in the U.S. from overdoses related to prescription opioids.(1,2) Scientific research has identified high-risk prescribing practices that have contributed to the opioid overdose epidemic, including overlapping opioid and benzodiazepine prescriptions.(3) Concurrent use of opioids and benzodiazepines, both central nervous system (CNS) depressants, increases the risk for severe respiratory depression, which can be fatal.(3,4)

According to the Centers for Disease Control and Prevention (CDC) Guideline for Prescribing Opioids for Chronic Pain – United States, 2016, clinicians should avoid concurrent prescribing of opioids and benzodiazepines whenever possible. (3) This is a Category A recommendation (applies to all persons; most patients should receive the recommended course of action) and is based on Type 3 evidence (observational studies or randomized clinical trials with notable limitations). In August 2016, the US Food and Drug Administration added concurrent use of opioids and benzodiazepines as a black box warning to prescription opioids (analgesic and cough medicine) and benzodiazepines. (4)

Several studies indicate that concurrent use of opioids and benzodiazepines puts patients at greater risk for a fatal overdose. Three studies of opioid overdose deaths found evidence of concurrent benzodiazepine use in 31%–61% of cases. (5-7) In the United States, the number of opioid overdose deaths involving benzodiazepines increased 14% on average for each year from 2006 through 2011. However, the number of opioid overdose deaths not involving benzodiazepines did not change significantly.(8) A case-cohort study found that concurrent use of benzodiazepines among US veterans raised the risk of drug overdose deaths four-fold (hazard ratio, 3.86, 95% confidence interval [CI] 3.49-4.26) compared with patients not using benzodiazepines.(9) In a large sample of privately insured patients from 2001-2013, opioid users who also used benzodiazepines were at substantially higher risk of an emergency department (ED) visit or hospital admission for opioid overdose (adjusted odds ratio 2.14; 95% CI, 2.05-2.24). If this association is causal, elimination of the

concurrent use could reduce the population risk of an ED visit or hospitalization for opioid overdose by 15%. (10)

Despite the risks, concurrent prescriptions for opioids and benzodiazepines are common and increasing. From 2001-2013, concurrent prescribing (overlap of at least one day) increased by nearly 80% (from 9% to 17%) among privately insured patients. (10) In one study, approximately half of the patients received both opioid and benzodiazepine prescriptions from the same prescriber on the same day.(11) In a 2015 analysis of Medicare Part D non-cancer and/or non-hospice patients on opioid therapy, the prevalence of benzodiazepine concurrent use was 24%.(12)

The PQA Concurrent Use of Opioids and Benzodiazepines measure evaluates a process that correlates with increased risk of opioid overdose. Efforts to prevent opioid overdose deaths should include a multi-faceted approach, including strategies that focus on monitoring and reducing opioid prescribing that has an unfavorable balance of benefit and harm for most patient populations. The measure excludes patients with cancer and those in hospice due to the unique therapeutic goals, ethical considerations, increased opportunities for medical supervision, and balance of risks and benefits with opioid therapy. (3)

1. Hedegaard H, Warner M, Miniño AM. Drug overdose deaths in the United States, 1999–2016. NCHS Data Brief, no 294. Hyattsville, MD: National Center for Health Statistics. 2017/ CDC. Wide-ranging online data for epidemiologic research (WONDER). Atlanta, GA: CDC, National Center for Health Statistics; 2016. Available at http://wonder.cdc.gov

2. Frenk SM, Porter KS, Paulozzi LJ. Prescription opioid analgesic use among adults: United States, 1999–2012. NCHS data brief, no 189. Hyattsville, MD: National Center for Health Statistics. 2015.

3. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016;65(1):1-49. doi:10.15585/mmwr.rr6501e1.

4. US Food and Drug Administration. FDA Drug Safety Communication: FDA warns about serious risks and death when combining opioid pain or cough medicines with benzodiazepines; requires its strongest warning. August 31, 2016. Available at: http://www.fda.gov/Drugs/DrugSafety/ucm518473.htm. Accessed: November 9, 2016.

5. Gomes T, Mamdani MM, Dhalla I a, Paterson JM, Juurlink DN. Opioid dose and drug-related mortality in patients with nonmalignant pain. Arch Intern Med. 2011;171(7):686-691. doi:10.1001/archinternmed.2011.117.

6. Dasgupta N, Funk MJ, Proescholdbell S, Hirsch A, Ribisl KM, Marshall S. Cohort Study of the Impact of High-dose Opioid Analgesics on Overdose Mortality. Pain Med. September 2015. doi:10.1111/pme.12907.

7. Jones CM, McAninch JK. Emergency Department Visits and Overdose Deaths from Combined Use of Opioids and Benzodiazepines. Am J Prev Med. 2015;49(4):493-501. doi:10.1016/j.amepre.2015.03.040.

8. Chen LH, Hedegaard H, Warner M. Drug-poisoning Deaths Involving Opioid Analgesics: United States, 1999-2011. NCHS Data Brief. 2014;(166):1-8.

9. Park TW, Saitz R, Ganoczy D, Ilgen MA, Bohnert ASB. Benzodiazepine prescribing patterns and deaths from drug overdose among US veterans receiving opioid analgesics?: case-cohort study. :1-8. doi:10.1136/bmj.h2698.

10. Sun EC, Dixit A, Humphreys K, et al. Association between concurrent use of prescription opioids and benzodiazepines and overdose: retrospective analysis. BMJ. 2017;356:j760. doi: 10.1136/bmj.j760. PMID: 28292769

11. Hwang CS, Kang EM, Kornegay CJ, Staffa JA, Jones CM, McAninch JK. Trends in the Concomitant Prescribing of Opioids and Benzodiazepines, 2002-2014. Am J Prev Med. 2016:1-10. doi:10.1016/j.amepre.2016.02.014.

12. CMS. Concurrent Use of Opioids and Benzodiazepines in a Medicare Part D Population. May 12, 2016. 2016. https://www.cms.gov/Medicare/Prescription-Drug-

Coverage/PrescriptionDrugCovContra/Downloads/Concurrent-Use-of-Opioids-and-Benzodiazepines-in-a-Medicare-Part-D-Population-CY-2015.pdf. Accessed December 6, 2016.

S.4. Numerator Statement: The number of individuals from the denominator with concurrent use of opioids and benzodiazepines for >=30 cumulative days during the measurement year.

S.6. Denominator Statement: The denominator includes individuals >=18 years of age with >=2 prescription claims for opioid medications on different dates of service and with >=15 cumulative days' supply during the measurement year. Individuals with cancer, sickle cell disease, or in hospice are excluded.

S.8. Denominator Exclusions: Individuals with cancer, sickle cell disease, or in hospice at any point during the measurement year are excluded from the denominator.

De.1. Measure Type: Process

S.17. Data Source: Claims, Enrollment Data

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Oct 26, 2018 **Most Recent Endorsement Date:** Oct 26, 2018

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

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

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

The developer provides the following evidence for this measure:

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

Summary of prior review in 2018

- The developer cited the CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016, which recommends clinicians should avoid prescribing opioid pain medication and benzodiazepines concurrently whenever possible (Recommendation Category: A; Evidence Type: 3).
- Category A recommendation: Applies to all persons; most patients should receive the recommended course of action. Type 3 evidence: Observational studies or randomized clinical trials with notable limitations.
- The developer cited four studies. Observational studies: a) three epidemiologic series of concurrent benzodiazepine use in large proportions of opioid-related overdose deaths, and b) one case-cohort study.

Changes to evidence from last review

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

☒ **The developer provided updated evidence for this measure:**

Updates:

- Updated evidence submitted includes four additional retrospective cohort studies, one case cohort study, and a technical brief from The Agency for Healthcare Research and Quality (AHRQ). The studies demonstrate the relationship between concurrent use of opioids and benzodiazepines and increased risk for overdose and other adverse events, as well as continued prevalence of concurrent use of opioids and benzodiazepines and room for improvement.

Questions for the Committee:

- The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?
- Questions specific to the measure information provided on evidence
    - What is the relationship of this measure to patient outcomes?
    - How strong is the evidence for this relationship?
    - Is the evidence directly applicable to the process of care being measured?

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) -> QQC Provided (Box 4) -> Quantity: High; Quality: Moderate; Consistency: Moderate (Box 5b) -> Moderate. The highest possible rating is High.

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

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

Maintenance measures – increased emphasis on gap and variation

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

Performance scores on the measure as specified are provided from implementations in the Medicare Part D Patient Safety Reports, Medicare Part D Display page, and the Medicaid Adult Core Set.

Medicare Part D Patient Safety Reports - Data are provided for the full 2018 data year and partial data from the 2019 data year reflecting the most recent made available to PQA at the time of submission (final

year-to-date rates are generated at the end of July of each year following the final prescription drug event (PDE) submission at the end of June, with the most recent 2019 data available generated from the April 2019 report, year-to-date).

Data are provided stratified by line of business (Medicare Advantage Prescription Drug Plan [MAPD], stand-alone Prescription Drug Plan [PDP]), inclusive of contracts with greater than 30 patients in the denominator.

- 2018 Data (MAPD n=605), Mean: 19.44%, St. Dev: 6.72%
- 2018 Data (PDP n=58), Mean: 19.36%, St. Dev: 4.78%
- 2019 Data (MAPD n=618), Mean: 17.39%, St. Dev: 6.15%
- 2019 Data (PDP n=57), Mean: 17.44%, St. Dev: 3.98%

Medicare Part D Display Page: The COB measure was implemented as a Part D Display measure in 2021 (using 2019 data). The Display data contain information for primarily the same entities over the same course of time as the 2019 Patient Safety data reported above, with a few minor differences. These differences include, as described above, that the Display data do not include contracts flagged as "Plan too new to be measured" or contracts that were otherwise not required to report the measure to CMS. Additionally, the Patient Safety data do not contain data from the full calendar year.

- For these reasons, performance distributions are expected to be similar, but not identical.

2019 data (MAPD N=479), Mean: 17.20%, St. Dev: 5.68%
2019 Data (PDP n=57), Mean: 17.43%, St. Dev: 3.98%

Medicaid Adult Core Set - Although the COB measure is not yet publicly reported in the Medicaid Adult Core Set program, preliminary deidentified data were obtained to support this NQF submission. These data include performance rates from 19 state Medicaid programs that reported on the measure for calendar year 2018, and one state that reported data from federal fiscal year 2018. Of these 20 states, 12 provided data on the Medicaid population only, 3 provided data on Medicaid and Dual Eligible populations, four provided data on Medicaid and CHIP populations, and 1 provided data on Medicaid, Dual Eligible, and CHIP population. The total measure denominator population across these 20 states was 765,514. Please note that these data did not undergo the program's final quality assurance and review associated with public reporting.

Performance distributions are provided below.

2018 data (N=20), Mean: 19.15%, St. Dev: 5.36%

Disparities

- The developer mentions that available performance data from measure implementation in the Medicare Part D Patient Safety Reports program, Medicare Part D Display Page, and Medicaid Adult Core Set did not include stratification by population groups. Therefore, the developer provides disparities data from original measure testing for the Medicare population.
  - The measure rate for the LIS group was higher than the rate for the non-LIS group.
  - Rates by age group trended down as age increased, and rates were higher among females than males.
    - Measure rates by LIS status (Medicare): LIS: 29.9%, Non-LIS: 19.9%
    - Measure rates by age band (Medicare): Ages 18-50: 37.2%, Ages 51-64: 33.8%, Ages 65-84: 19.4%, Ages 85+: 16.7%
    - Measure rates by sex (Medicare): Male: 21.3%, Female: 26.5%
  - For the original Medicaid testing described in the testing form, the developer reports data for age bands and by sex:
    - Measure rates by age band (Medicaid): Ages 18-50: 3.1%, Ages 51-64: 6.4%, Ages 65-84: 1.3%, Ages 85+: 0.6%
    - Measure rates by sex (Medicaid): Male: 3.5%, Female: 3.7%

The developer also provides supporting literature that are consistent with performance disparities identified in testing, with evidence for additional potential disparities by race, disability, and dual-eligibility, among others.

Questions for the Committee:

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

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

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

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures–are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- Evidence is appropriate and current
- Evidence to support the measure is strong and based on empirical data and/or established guidelines. No concerns
- Strong evidence to support concurrent prescription of opioids and benzodiazepines has a high correlation with overdose and mortality.
- high
- We do not know the outcomes, which are likely to depend on the dose of opioids taken. This substantially weakens the link between this process measure and outcomes.
- high level of evidence
- Reasonable evidence
- process 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?

- There is a performance gap. Disparities by LIS, age, etc. are presented.
- Significant spread in the data indicating opportunities for improvement. Disparities noted in care that could rise to clinically meaningful differences by age and sex.
- Not clear from evidence what the current prevalence of this co-prescribing is. Data used to support is several years old.
- High - existing performance gap.
- There are gaps, but as a national performance measure this process measure falls short.
- high
- Clear gaps
- High evidence presented, current Medicare data does not allow for stratification, developer used previous data, ranked high for performance gaps

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population at 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**

Reliability

- The developer conducted measure score reliability testing on data from the 2018 Part D Patient Safety Reports using the Adams beta-binomial methodology.
- Estimates were only computed for contracts with greater than 30 patients in the denominator.
- Table 4A provides the distribution of reliability estimates by line of business.

Table 4A. Plan Reliability Score Distribution for the Part D Patient Safety Reports

Statistic	Values (MAPD)	Values (PDP)
10th Percentile	.53	.72
25th Percentile	.79	.89
Median	.95	.98
75th Percentile	.99	.996
90th Percentile	.995	.999
Mean	**.86**	**.91**
Standard Deviation	**.18**	**.15**

Validity

- The developer conducted measure score criterion validity testing. The developer evaluated the correlation between plan-level performance on the COB measure as specified and plan-level rates of a composite of inpatient stays and emergency department utilization due to opioid- and benzodiazepine-related adverse events (OBRAEs).
- The developer hypothesized an expected convergent relationship between measure rates and OBRAEs; the better a given plan performs on the COB measure (i.e., lower rate), the lower plan-level rates of OBRAEs are hypothesized to be.
- The developer reported that within the Medicare 5% sample, the Spearman's correlation coefficient was .45 within PDPs (moderate) [p<.0001] and .21 for MAPDs (weak) [p=.001].

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?

Questions for the Committee regarding validity:

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

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

Specifications precise, unambiguous, and complete (Box 1) -> Empirical reliability testing conducted using statistical tests with the measure as specified (Box 2) -> Reliability testing conducted with computed performance measure scores for each measured entity (Box 4) -> Method described (signal-to-noise analysis) and appropriate for assessing the proportion of variability due to real differences among measured entities (Box 5) -> Moderate certainty or confidence that the performance measure scores are reliable (distribution of percentiles to mean) (Box 6b) – Moderate. The highest possible rating is High.

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

All relevant potential threats to validity empirically assessed (Box 1) -> Empirical validity testing conducted using the measure as specified and appropriate statistical test (Box 2) -> Validity testing conducted with computed performance measure scores for each measured entity (Box 5) -> Method described appropriate for assessing conceptually and theoretically sound hypothesized relationships (Box 6) -> Moderate certainty or confidence that the performance measure scores are a valid indicator of quality (Box 7b) – Moderate. The highest possible rating is High.

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?

- no concerns
- Moderate reliability but data appears robust. Medicare plans slightly lower reliability, but looks related to sample size. Signal to noise ratios good
- Not clear if data from Medicare advantage (managed medicare) is included?
- moderate reliability. definition question - why sickle cell is excluded - just because it's acceptable to use chronic opioids for pain in sickle cell patients does not mean that concurrent use of benzos is acceptable. Granted this is a very small proportion of the population.
- Measure specifications are inadequate because they do not account for dose administered.
- moderate level prelim rating
- adequate
- not reviewed by panel - moderate reliability

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure; reliability testing and results for the measure?

- no concerns
- No
- same concern as above
- No concerns
- as stated above
- no concerns
- No
- moderate reliability

2b1. Validity -Testing: Do you have any concems with the validity testing and results for the measure?

- no concerns
- No. While correlations were lower than anticipated, the developers point out expected noise in ED/hospital admission rates that explains low noise. The measure has strong face validity and may be difficult to measure full battery of adverse events correlated.
- no
- No concerns, agree with "moderate". some plans' correlation with opioids/benzo related adverse events is rather low at 0.21 but not sure how high a correlation one would expect.
- To be truly valid, this should be converted to an outcome measure
- no concerns
- no
- moderate validity

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?

- no concerns
- No
- missing data constitutes a threat to the validity of this measure.
- No concerns.

- none apparent
- no concerns
- none
- moderate validity

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?

- no risk adjustment
- Exclusions are appropriate. However, consideration of exclusions for long term care residents and palliative care patients may minimize risk of falls for this population which would not be aligned with therapeutic goals for either
- Significant variation was shown to be statistically relevant for both Medicare and Medicare plans that are significant and clinically meaningful.
- no discussion of risk adjustment - I believe this should at least be discussed even if developer chooses to not risk adjust.
- na
- no concerns
- no issues
- no risk adjustment

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3389

Measure Title: Concurrent Use of Opioids and Benzodiazepines (COB)

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.**
   - No concerns.

RELIABILITY: TESTING

Type of measure:

☐ **Outcome (including PRO-PM)** ☐ **Intermediate Clinical Outcome** ☒ **Process**

☐ **Structure** ☐ **Composite** ☐ **Cost/Resource Use** ☐ **Efficiency**

Data Source:

☐ **Abstracted from Paper Records** ☒ **Claims** ☐ **Registry**
☐ **Abstracted from Electronic Health Record (EHR)** ☐ **eMeasure (HQMF) implemented in EHRs**
☐ **Instrument-Based Data** ☒ **Enrollment Data** ☐ **Other (please specify)**

Level of Analysis:

☐ **Individual Clinician** ☐ **Group/Practice** ☐ **Hospital/Facility/Agency** ☒ **Health Plan**
☐ **Population: Regional, State, Community, County or City** ☐ **Accountable Care Organization**
☐ **Integrated Delivery System** ☐ **Other (please specify)**

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

3) Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Neither**

4) Reliability testing was conducted with the data source and level of analysis indicated for this measure
☒ **Yes** ☐ **No**

5) If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of **patient-level data** conducted?
☐ **Yes** ☐ **No**

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

Submission document: Testing attachment, section 2a2.2

- The reliability of the computed measure scores was measured as the ratio of signal-to-noise.
- The signal is the proportion of the variability in measured performance that can be explained by true differences in plan (or contract) performance.
- Reliability scores range from 0 to 1, with a score of 0 signifying that all variation is due to measurement error.
- A value of 1 signifies that the variation represents true differences in performance scores between plans.
- A beta-binomial model was used to calculate plan-specific reliability scores.
- The reliability score is defined as the ratio of the plan-to-plan variance to the sum of the plan-to-plan variance and the plan-specific error.
- The plan-to-plan variance is an estimate of the variance of the true rates.
- The plan-specific error variance is the sampling or measurement error.

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

Submission document: Testing attachment, section 2a2.3

- In order to demonstrate reliability in the COB measure's implementation in the field, PQA conducted reliability analyses on data from the 2018 Part D Patient Safety Reports using the Adams beta-binomial methodology.
- Estimates were only computed for contracts with greater than 30 patients in the denominator.
- Table 4A provides the distribution of reliability estimates by line of business.

Table 4A. Plan Reliability Score Distribution for the Part D Patient Safety Reports

Statistic	Values (MAPD)	Values (PDP)
10th Percentile	.53	.72
25th Percentile	.79	.89
Median	.95	.98
75th Percentile	.99	.996
90th Percentile	.995	.999
Mean	**.86**	**.91**
Standard Deviation	**.18**	**.15**

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)

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

- Guidance from Reliability Algorithm:
- Specifications precise, unambiguous, and complete (Box 1) -> Empirical reliability testing conducted using statistical tests with the measure as specified (Box 2) -> Reliability testing conducted with computed performance measure scores for each measured entity (Box 4) -> Method described (signal-to-noise analysis) and appropriate for assessing the proportion of variability due to real differences among measured entities (Box 5) -> Moderate certainty or confidence that the performance measure scores are reliable (distribution of percentiles to mean) (Box 6b) – Moderate. The highest possible rating is High.

VALIDITY: TESTING

12) Validity testing level: ☒ Measure score ☐ Data element ☐ Both

13) **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**

NOTE that data element validation from the literature is acceptable.

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

☒ **Yes**

☐ **No**

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

14) Method of establishing validity of the measure score:

☐ **Face validity**

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

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

15) 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)

16) **Assess the method(s) for establishing validity**

Submission document: Testing attachment, section 2b2.2

- The developer conducted measure score criterion validity testing. The developer evaluated the correlation between plan-level performance on the COB measure as specified and plan-level rates of a composite of inpatient stays and emergency department utilization due to opioid- and benzodiazepine-related adverse events (OBRAEs).
- The developer hypothesized an expected convergent relationship between measure rates and OBRAEs; the better a given plan performs on the COB measure (i.e., lower rate), the lower plan-level rates of OBRAEs are hypothesized to be.

17) Assess the results(s) for establishing validity

Submission document: Testing attachment, section 2b2.3

- The developer reported that within the Medicare 5% sample, the Spearman's correlation coefficient was .45 within PDPs (moderate) [p<.0001] and .21 for MAPDs (weak) [p=.001].

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

18) **Please describe any concerns you have with measure exclusions.**

Submission document: Testing attachment, section 2b2.

- No concerns.

19) Risk Adjustment

Submission Document: Testing attachment, section 2b3

19a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

19c. **Social risk adjustment:**

19c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☒ Not applicable

19c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

19d. **Risk adjustment summary:**

19d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

19d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

19d.5. Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

19e**. Assess the risk-adjustment approach**

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

Submission document: Testing attachment, section 2b4.

- To assess significant differences in measure rates, the data described in sections 1.5 and 1.6 were used to calculate the mean, median, standard deviation, and interquartile range for the measure rates for the Medicare and Medicaid (MAX) populations.
- In addition, the rates were divided into quartiles, and a Student's t-test was used to compare the rates of the plans in the 25th percentile to the rates of the plans in the 75th percentile.
- The mean rate for the Medicare population was 22.2%, with a median rate of 21.4%, with the lowest plan contract rate at 2.1% and the highest plan contract rate of 44.7%.
- The mean rate for the Medicaid MAX population was 3.8%, with a median rate of 2.9%. The lowest plan contract rate was 0.0% and the highest plan contract rate was 18.7%.
- For the Medicare population, the measure rates showed significant variation, with a standard deviation of 7.3% and an Interquartile Range of 9.9%. There is a statistically significant difference in measure rates between the top and bottom quartile of the plans included in the testing ($P < .0001$ at alpha = 0.05). This variation shows that there are statistically significant and clinically meaningful differences in rates across plans.
- For the Medicaid population, the measure rates showed significant variation, with a standard deviation of 3.2% and an Interquartile Range of 3.4%. There is a statistically significant difference in measure rates between the top and bottom quartile of the plans included in the testing ($P < .0001$ at alpha = 0.05). This variation shows that there are statistically significant and clinically meaningful differences in rates across plans.

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

Submission document: Testing attachment, section 2b5.

N/A

22) Please describe any concerns you have regarding missing data.

Submission document: Testing attachment, section 2b6.

No missing data was found in the testing of this measure.

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

All relevant potential threats to validity empirically assessed (Box 1) -> Empirical validity testing conducted using the measure as specified and appropriate statistical test (Box 2) -> Validity testing conducted with computed performance measure scores for each measured entity (Box 5) -> Method described appropriate for assessing conceptually and theoretically sound hypothesized relationships (Box 6) ->

Moderate certainty or confidence that the performance measure scores are a valid indicator of quality (Box 7b) – Moderate. The highest possible rating is High.

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.

- Medical claims data, Prescription claims data, Enrollment Data
- ALL data elements are in defined fields in electronic claims.
- PQA does not intend to develop an electronic clinical quality measure (eCQM) version of this health plan claims-based performance measure. However, PQA is currently exploring opportunities to convert existing claims-based measures to a digital clinical quality measure (dCQM) format, to align with CMS' stated goal of using dCQMs by 2025.
- PQA is not aware of difficulties in implementing the measure into the programs described in 4.1. The measure is specified using prescription and medical claims data, which are readily available and accurate.
- All uses of PQA Measures are subject to such conditions as PQA specifies, and will be subject to a license agreement specifying the terms of use and the license fee. Government agencies do not pay a license royalty.

Questions for the Committee:

Are the required data elements routinely generated and used during care delivery?

Are the required data elements available in electronic form, e.g., EHR or other electronic sources?

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?

- no concerns
- Extractions from plan prescription data; no concerns
- the data should be readily available based on prescribing information
- No concerns.
- none
- moderate
- No concerns
- moderate feasibility - combination of claims and enrollment data

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

Medicare Part D Display page

The developer states that CMS will consider this measure for the 2023 Star Ratings (using 2021 data) pending rulemaking.

Medicare Part D Patient Safety Reports

Centers for Medicare & Medicaid Services (CMS) Medicaid Adult Core Set.

National program with state-level voluntary reporting.

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

Since initial development and endorsement, the developer received feedback that an exclusion for patients with a sickle cell disease diagnosis is appropriate for the measure from a variety of sources. After soliciting expert input and completing review, these recommendations were presented to and approved by PQA's Measure Update Panel, and PQA's Quality Metrics Expert Panel. Due to these considerations, and their unique therapeutic goals, ethical considerations, opportunities for medical supervision, and balance of risks and benefits, individuals with a diagnosis of sickle cell disease are excluded from this measure.

The developer received feedback from measure users suggesting that a palliative care and long-term care exclusions may be appropriate for the measure. As a result, the developer is evaluating the appropriateness of these exclusions for future updates to the measure.

Additional Feedback:

N/A

***Questions for the Committee**:*

How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?

How has the measure been vetted in real-world settings by those being measured or others?

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

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

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

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

Improvement results

- Data from 2018 and 2019 in the Medicare Part D Patient Safety Reports demonstrate a downward trend across both the MAPD and PDP lines of business. In addition, the performance distributions demonstrate variation and room for improvement.
- Data from the Medicare Part D Display page were only available for calendar year 2019, which does not allow for trend analysis. However, given that the measured population is nearly identical to the Medicare Part D Patient Safety reports as noted in section 1b.2, those data can be referenced to gain insight in COB measure rate trends in the Medicare Part D Display page.
- Data for the Medicaid Adult Core Set were only available for calendar year 2018, which does not allow for trend analysis. As more states report on the measure, PQA will continue to monitor trends for improvement.

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

Unexpected findings (positive or negative) during implementation

The developer does not report any unexpected findings.

Potential harms

N/A

Additional Feedback:

N/A

***Questions for the Committee**:*

How can the performance results be used to further the goal of high-quality, efficient healthcare?

Do the benefits of the measure outweigh any potential unintended consequences?

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

Committee Pre-evaluation Comments:

Criteria 4: Usability and Use

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

- publicly reported; being considered for accountability for Medicare Part D plans
- Publicly reported. No concerns.
- The information is publicly reported but not currently used in an accountability program, although the developer states CMS will consider using for star ratings starting in 2023. This is hard to rely on given the PHE.
- No concerns. Rating is "pass".
- Further work on exclusions are ongoing - long term care and palliative care
- high prelim rating
- Reported and planned to be used
- pass for use

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.

- no unintended consequences evident
- No reports of unintended consequences.
- Highly usable to decrease patient risk.
- No known unintended consequences.
- Without outcome data, one cannot judge unintended consequences.
- benefits > harms
- Usable
- High usability

Criterion 5: Related and Competing Measures

Related or competing measures

Related Measures:

- 2940: Use of Opioids at High Dosage in Persons Without Cancer
- 2950: Use of Opioids from Multiple Providers in Persons Without Cancer
- 2951: Use of Opioids from Multiple Providers and at High Dosage in Persons Without Cancer
- 3316: Safe Use of Opioids – Concurrent Prescribing
- 3541: Annual Monitoring for Persons on Long-Term Opioid Therapy (AMO)
- 3558: Initial Opioid Prescribing for Long Duration (IOP-LD)
- Use of Opioids at High Dosage (NCQA)
- Use of Opioids from Multiple Providers (NCQA)

Competing Measures:

- There are no competing measures (i.e., those that addresses both the same measure focus and the same target population).

Harmonization

At time of maintenance, PQA has also identified the 3316e: Safe Use of Opioids – Concurrent Prescribing measure as related. Although the area of focus overlaps, 3316e is specified at the facility level as an eCQM, as opposed to 3389, which is specified at the health plan level and is claims-based.

PQA identified the 3558: Initial Opioid Prescribing for Long Duration and 3541: Annual Monitoring for Persons on Long-Term Opioid Therapy) measures as related to opioid prescribing, although the areas of focus (initial opioid prescribing and annual monitoring) are different than 3389 (concurrent use of opioids and benzodiazepines).

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?

- no competing measures
- no other measures evaluating concurrent use; other overlaps are appropriate
- no
- No concerns.
- no
- several, listed in worksheet
- no issues
- multiple related measures with varying definitions and measure characteristics

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 06/03/2021

Comment by: American Medical Association

The American Medical Association (AMA) appreciates the opportunity to comment on Measure #3389, Concurrent Use of Opioids and Benzodiazepines. While we appreciate the updates made to the measure including the addition of an exclusion for sickle cell disease, we continue to believe that the measure lacks the precision needed to ensure that only those patients for whom concurrent prescribing of two or more opioids or an opioid and benzodiazepine are included in the denominator. The patient population could likely include patients for whom concurrent prescribing of these medications may be appropriate, particularly those with chronic pain.

In addition and more importantly, the National Quality Forum (NQF) and the measure developer must consider the potential for unintended consequences and complete robust evaluations to minimize these risks. In fact, we believe that the narrow and reactionary response to the drug overdose epidemic has exacerbated the stigma around opioid use and made it more difficult for patients with pain or opioid use disorder to receive treatment. Research continues to demonstrate that individuals may or may not have access to pain management therapies based on their race/ethnicity and measures that may further exacerbate this problem should be avoided (Goshal, 2020). In addition to stigmatization of those with substance use disorder, patients with other complex pain management conditions (such as sickle cell disease) are often viewed as opioid-seeking when presenting in the emergency department. Therefore, we urge NQF to consider whether this and other measures that are focused on areas such as opioid dose and duration continue to be appropriate.

As a result, the AMA believes that there is a significant risk for performance to be inaccurately represented. More importantly, there is a substantial risk that patients for whom these medications may be warranted will not receive appropriate therapies, leading to potential adverse outcomes, including depression, loss of function and other negative unintended consequences.

The AMA believes that quality measurement needs to focus on how well patients' pain is controlled, whether functional improvement goals are met, and what therapies are being used to manage pain. If pain can be well controlled and function improved without the need of these concurrent medications, then that is an indication of good patient care but the measure must precisely define the patients for which it is appropriate and be tested to ensure that negative unintended consequences are not experienced by patients. We do not believe that this measure as specified is an appropriate goal as it may leave patients without access to needed therapies.

The AMA supports addressing the opioid crisis through quality measurement in addition to other avenues but strongly believes that any measure endorsed by NQF must also demonstrate that it does not compromise patient care. As a result, the AMA does not support continued endorsement of measure #3389.

Reference:

Goshal M, Shapiro H, Todd, K, Schatman ME. Chronic noncancer pain management and systemic racism: Time to move toward equal care standards. J Pain Res. 2020;13:2825-2836.

Of the 1 NQF member who have submitted a support/non-support choice:

1 supports the measure

0 do not support the measure

Developer Submission

NQF #: 3389

Corresponding Measures:

De.2. Measure Title: Concurrent Use of Opioids and Benzodiazepines (COB)

Co.1.1. Measure Steward: PQA, Inc.

De.3. Brief Description of Measure: The percentage of individuals >=18 years of age with concurrent use of prescription opioids and benzodiazepines during the measurement year.

A lower rate indicates better performance.

1b.1. Developer Rationale: Overdose deaths involving prescription opioids were five times higher in 2016 than in 1999, and more than 200,000 people have died in the U.S. from overdoses related to prescription opioids.(1,2) Scientific research has identified high-risk prescribing practices that have contributed to the opioid overdose epidemic, including overlapping opioid and benzodiazepine prescriptions.(3) Concurrent use of opioids and benzodiazepines, both central nervous system (CNS) depressants, increases the risk for severe respiratory depression, which can be fatal.(3,4)

According to the Centers for Disease Control and Prevention (CDC) Guideline for Prescribing Opioids for Chronic Pain – United States, 2016, clinicians should avoid concurrent prescribing of opioids and benzodiazepines whenever possible.(3) This is a Category A recommendation (applies to all persons; most patients should receive the recommended course of action) and is based on Type 3 evidence (observational studies or randomized clinical trials with notable limitations). In August 2016, the US Food and Drug Administration added concurrent use of opioids and benzodiazepines as a black box warning to prescription opioids (analgesic and cough medicine) and benzodiazepines.(4)

Several studies indicate that concurrent use of opioids and benzodiazepines puts patients at greater risk for a fatal overdose. Three studies of opioid overdose deaths found evidence of concurrent benzodiazepine use in 31%–61% of cases.(5-7) In the United States, the number of opioid overdose deaths involving benzodiazepines increased 14% on average for each year from 2006 through 2011. However, the number of opioid overdose deaths not involving benzodiazepines did not change significantly.(8) A case-cohort study found that concurrent use of benzodiazepines among US veterans raised the risk of drug overdose deaths four-fold (hazard ratio, 3.86, 95% confidence interval [CI] 3.49-4.26) compared with patients not using benzodiazepines.(9) In a large sample of privately insured patients from 2001-2013, opioid users who also used benzodiazepines were at substantially higher risk of an emergency department (ED) visit or hospital admission for opioid overdose (adjusted odds ratio 2.14; 95% CI, 2.05-2.24). If this association is causal, elimination of the concurrent use could reduce the population risk of an ED visit or hospitalization for opioid overdose by 15%.(10)

Despite the risks, concurrent prescriptions for opioids and benzodiazepines are common and increasing. From 2001-2013, concurrent prescribing (overlap of at least one day) increased by nearly 80% (from 9% to 17%) among privately insured patients.(10) In one study, approximately half of the patients received both opioid and benzodiazepine prescriptions from the same prescriber on the same day.(11) In a 2015 analysis of Medicare Part D non-cancer and/or non-hospice patients on opioid therapy, the prevalence of benzodiazepine concurrent use was 24%.(12)

The PQA Concurrent Use of Opioids and Benzodiazepines measure evaluates a process that correlates with increased risk of opioid overdose. Efforts to prevent opioid overdose deaths should include a multi-faceted approach, including strategies that focus on monitoring and reducing opioid prescribing that has an unfavorable balance of benefit and harm for most patient populations. The measure excludes patients with cancer and those in hospice due to the unique therapeutic goals, ethical considerations, increased opportunities for medical supervision, and balance of risks and benefits with opioid therapy.(3)

1.Hedegaard H, Warner M, Miniño AM. Drug overdose deaths in the United States, 1999–2016. NCHS Data Brief, no 294. Hyattsville, MD: National Center for Health Statistics. 2017/ CDC. Wide-ranging online data for epidemiologic research (WONDER). Atlanta, GA: CDC, National Center for Health Statistics; 2016. Available at http://wonder.cdc.gov

2. Frenk SM, Porter KS, Paulozzi LJ. Prescription opioid analgesic use among adults: United States, 1999–2012. NCHS data brief, no 189. Hyattsville, MD: National Center for Health Statistics. 2015.

3. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016;65(1):1-49. doi:10.15585/mmwr.rr6501e1.

4. US Food and Drug Administration. FDA Drug Safety Communication: FDA warns about serious risks and death when combining opioid pain or cough medicines with benzodiazepines; requires its strongest warning. August 31, 2016. Available at: http://www.fda.gov/Drugs/DrugSafety/ucm518473.htm. Accessed: November 9, 2016.

5. Gomes T, Mamdani MM, Dhalla I a, Paterson JM, Juurlink DN. Opioid dose and drug-related mortality in patients with nonmalignant pain. Arch Intern Med. 2011;171(7):686-691. doi:10.1001/archinternmed.2011.117.

6. Dasgupta N, Funk MJ, Proescholdbell S, Hirsch A, Ribisl KM, Marshall S. Cohort Study of the Impact of High-dose Opioid Analgesics on Overdose Mortality. Pain Med. September 2015. doi:10.1111/pme.12907.

7. Jones CM, McAninch JK. Emergency Department Visits and Overdose Deaths from Combined Use of Opioids and Benzodiazepines. Am J Prev Med. 2015;49(4):493-501. doi:10.1016/j.amepre.2015.03.040.

8. Chen LH, Hedegaard H, Warner M. Drug-poisoning Deaths Involving Opioid Analgesics: United States, 1999-2011. NCHS Data Brief. 2014;(166):1-8.

9. Park TW, Saitz R, Ganoczy D, Ilgen MA, Bohnert ASB. Benzodiazepine prescribing patterns and deaths from drug overdose among US veterans receiving opioid analgesics?: case-cohort study. :1-8. doi:10.1136/bmj.h2698.

10. Sun EC, Dixit A, Humphreys K, et al. Association between concurrent use of prescription opioids and benzodiazepines and overdose: retrospective analysis. BMJ. 2017;356:j760. doi: 10.1136/bmj.j760. PMID: 28292769

11. Hwang CS, Kang EM, Kornegay CJ, Staffa JA, Jones CM, McAninch JK. Trends in the Concomitant Prescribing of Opioids and Benzodiazepines, 2002-2014. Am J Prev Med. 2016:1-10. doi:10.1016/j.amepre.2016.02.014.

12. CMS. Concurrent Use of Opioids and Benzodiazepines in a Medicare Part D Population. May 12, 2016. 2016. https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Concurrent-Use-of-Opioids-and-Benzodiazepines-in-a-Medicare-Part-D-Population-CY-2015.pdf. Accessed December 6, 2016.

S.4. Numerator Statement: The number of individuals from the denominator with concurrent use of opioids and benzodiazepines for >=30 cumulative days during the measurement year.

S.6. Denominator Statement: The denominator includes individuals >=18 years of age with >=2 prescription claims for opioid medications on different dates of service and with >=15 cumulative days' supply during the measurement year. Individuals with cancer, sickle cell disease, or in hospice are excluded.

S.8. Denominator Exclusions: Individuals with cancer, sickle cell disease, or in hospice at any point during the measurement year are excluded from the denominator.

De.1. Measure Type: Process

S.17. Data Source: Claims, Enrollment Data

S.20. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: Oct 26, 2018 **Most Recent Endorsement Date:** Oct 26, 2018

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

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

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

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

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

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

COB_2021Maintenance_Evidence_FNL-637522777157744034.docx

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

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

Yes

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 3389

Measure Title: Concurrent Use of Opioids and Benzodiazepines

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

Date of Submission: **4/1/2021**

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

Outcome

☐ Outcome:

☐ Patient-reported outcome (PRO):

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

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

☒ Process: **Concurrent use of opioid medications and benzodiazepine medications**

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

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

The measured process, concurrent use of opioids and benzodiazepines, correlates with negative health outcomes. Scientific research has identified high-risk prescribing practices that have contributed to the opioid overdose epidemic, including overlapping opioid and benzodiazepine prescriptions. The *Centers for Disease Control and Prevention (CDC) Guideline for Prescribing Opioids for Chronic Pain – United States, 2016*, provides a category A recommendation (applies to all persons; most patients should receive the recommended course of action) that prescribers should avoid concurrent prescriptions of opioids and benzodiazepines. Opioids and benzodiazepines are both central nervous system (CNS) depressants and can increase the risk for severe respiratory depression and fatal overdose. Few medication situations warrant concurrent use of opioids and benzodiazepines, specifically oncology, sickle cell disease, and hospice, which are excluded from the measure. The lack of a therapeutic benefit combined with increased risk for overdose is the rationale to support this process measure.

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

N/A

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

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

N/A

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Systematic Review	Evidence
Source of Systematic Review: Title Author Date Citation, including page number URL	CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. Dowell D, Haegerich TM, Chou R. March 18, 2016 MMWR Recomm Rep. 2016 Mar 18;65(1):1-49. doi: 10.15585/mmwr.rr6501e1. Available at: http://www.cdc.gov/drugoverdose/prescribing/guideline.html Also, the associated Clinical Evidence Review (http://stacks.cdc.gov/view/cdc/38026), and Contextual Evidence Review (http://stacks.cdc.gov/view/cdc/38027).
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.	CDC Guideline: Recommendation 11, pages 31-32, "Clinicians should avoid prescribing opioid pain medication and benzodiazepines concurrently whenever possible (recommendation category: A, evidence type: 3)."
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	CDC Guideline: Type 3 evidence: Observational studies or randomized clinical trials with notable limitations.
Provide all other grades and definitions from the evidence grading system	CDC Guideline: Evidence Type: Based on study design as well as a function of limitations in study design or implementation, imprecision of estimates, variability in findings, indirectness of evidence, publication bias, magnitude of treatment effects, dose-response gradient, and constellation of plausible biases that could change effects. Type 1 evidence: Randomized clinical trials or overwhelming evidence from observational studies. Type 2 evidence: Randomized clinical trials with important limitations, or exceptionally strong evidence from observational studies. Type 3 evidence: Observational studies or randomized clinical trials with notable limitations. Type 4 evidence: Clinical experience and observations, observational studies with important limitations, or randomized clinical trials with several major limitations.
Grade assigned to the **recommendation** with definition of the grade	CDC Guideline: Category A recommendation: Applies to all persons; most patients should receive the recommended course of action.
Provide all other grades and definitions from the recommendation grading system	CDC Guideline: Recommendation Categories Based on evidence type, balance between desirable and undesirable effects, values and preferences, and resource

Systematic Review	Evidence
	allocation (cost). Category A recommendation: Applies to all persons; most patients should receive the recommended course of action. Category B recommendation: Individual decision making needed; different choices will be appropriate for different patients. Clinicians help patients arrive at a decision consistent with patient values and preferences and specific clinical situations.
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	• Quantity: four studies • Quality: Observational studies; a) three epidemiologic series of concurrent benzodiazepine use in large proportions of opioid-related overdose deaths, and b) one case-cohort study.
Estimates of benefit and consistency across studies	Not provided.
What harms were identified?	The Clinical Evidence Review did not address risks of benzodiazepine co-prescription among patients prescribed opioids. However, the Contextual Evidence Review found evidence in epidemiologic series of concurrent benzodiazepine use in large proportions of opioid-related overdose deaths, and a case-cohort study found concurrent benzodiazepine prescription with opioid prescription to be associated with a near quadrupling of risk for overdose death compared with opioid prescription alone.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	1. Sun EC, Dixit A, Humphreys K, et al. Association between concurrent use of prescription opioids and benzodiazepines and overdose: retrospective analysis. BMJ. 2017;356:j760. doi: 10.1136/bmj.j760. PMID: 28292769 2. Gaither JR, Goulet JL, Becker WC, et al. The Association Between Receipt of Guideline-Concordant Long-Term Opioid Therapy and All-Cause Mortality. J Gen Intern Med 2016; 31:492 3. Dasgupta N, Funk MJ, Proescholdbell S, et al. Cohort Study of the Impact of High-Dose Opioid Analgesics on Overdose Mortality. Pain Med 2016; 17:85. The studies listed above do not change the conclusion from the SR. All support that the measured process correlates with negative health outcomes.

1a.4 OTHER SOURCE OF EVIDENCE

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

N/A

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

In a retrospective observational study (N=315,428), Sun et al. reported that opioid users who also used benzodiazepines were at higher risk of an emergency department visit or hospital admission for opioid overdose (adjusted odds ratio [aOR] 2.14; 95% Confidence Interval [CI], 2.05-2.24). (1) The authors estimated that the elimination of the concurrent use of opioids and benzodiazepines could reduce the population risk of an emergency department visit or hospital admission for opioid overdose by 15%.

In a retrospective observational study (N=17,044), Gaither et al. evaluated the association between receipt of guideline-concordant long-term opioid therapy (>90 days) among HIV-infected patients with 1-year all-cause mortality. (2) Patients prescribed benzodiazepines concurrent with opioids, defined as pharmacy documentation that the patient was prescribed a benzodiazepine greater than 7 days between start date and end of 180 days of long-term opioid therapy, had a higher risk of mortality (matched hazard ratio [HR] 1.39; 95% CI, 1.12-1.66).

In a prospective observational cohort study with one year of follow-up (N=2,182,374 with opioid prescriptions), Dasgupta et al. observed that rates of overdose death among patients on concurrent opioids and benzodiazepines in North Carolina were ten times higher (7 per 10,000 person-years; 95% CI 6.3-7.8) than opioid monotherapy (0.7 per 10,000 person-years; 95% CI 0.6-0.9). (3)

In August 2016, the US Food and Drug Administration (FDA) added Boxed Warnings to prescription drug labeling for prescription opioid pain and prescription opioid cough medications, and benzodiazepines, based on a review of the literature that found that combined use of opioids with benzodiazepines or other drugs that depress the central nervous system (CNS) has resulted in serious side effects, including slowed or difficult breathing and deaths. (4)

---UPDATED FOR MAINTENANCE---

Additional evidence further demonstrates the relationship between concurrent use of opioids and benzodiazepines and increased risk for overdose and other adverse events, as well as demonstrates continued prevalence of concurrent use of opioids and benzodiazepines and room for improvement.

A case-cohort study examined the association between benzodiazepine prescribing patterns and risk for drug overdose among US veterans (N=420,386) receiving opioids between 2004 and 2009. (5) During the study period, approximately 27% of veterans who received opioid analgesics also received benzodiazepines. Risk of death from drug overdose among those receiving opioids was substantially increased for those with current benzodiazepine prescriptions compared to those with no benzodiazepine prescription (Adjusted HR 3.86; 95% CI 3.49-4.26). Risk of death from drug overdose increased as daily benzodiazepine dose increased: compared to a reference dose of >0-10 mg/day, hazard ratios were >10-20 mg/day (1.69; 95% CI 1.42-2.01), >20-30 mg/day (2.34; 95% CI 1.91-2.86), >30-40 mg/day (2.65; 95% CI 2.10-3.33), >40 mg/day (3.06; 95% CI 2.38-3.92).

A retrospective cohort study examined patterns of opioid use among Medicaid beneficiaries in Washington state and their associations with opioid-related mortality from 2006 to 2010. (6) The study focused on noncancer patients with at least one opioid prescription (N=150,821). Risk was particularly high for opioids combined with benzodiazepines and skeletal muscle relaxants (adjusted HR 12.6; 95% CI 8.9–17.9). Even at low opioid doses, patients using sedative-hypnotics concurrently had 5.6 times the risk than patients without sedative-hypnotics (adjusted HR 5.6; 95% CI 1.6–19.3).

A retrospective cohort study of Medicare Part D claims data from 2013-2014 examined risk of opioid-related overdose in beneficiaries who filled at least one prescription for an opioid (N=71,248). (7) The study sought to evaluate the exposure-response association between days with concurrent use and risk of overdose. The study found that risk of overdose was highest in the early phase of concurrent use: The hazard ratio (compared to no opioid use) for overdose during the first 90 days was 5.05 (95% CI 3.68-6.93), compared to 1.87 (95% CI 1.25-2.80) for days 91 to 180, among those who did not have an event before 90 days. The authors concluded that policies deterring concurrent opioid and benzodiazepine use is warranted.

A retrospective cohort study of 245,954 non-cancer, disabled Medicare beneficiaries with >2 opioid prescriptions using data from 2013 through 2015 examined the association between concurrent opioid and benzodiazepine use and subsequent overdose risk. (8) The study used PQA's measure specifications to define concurrent use of opioids and benzodiazepines. Concurrent use of opioids and benzodiazepines remained stable across years at ~34% among the study population, and exposure to concurrent use of opioids and benzodiazepines was associated with a significant increase in overdose risk the for subsequent year (HR 1.82; 95% CI 1.58-2.10).

A retrospective cohort study of 2013-2014 Medical Expenditure Panel Survey data (N=16,815 survey-weighted to represent 321 million lives in the US) evaluated the risk of emergency department use associated with "double threat" concurrent prescribing of opioids and benzodiazepines. (9) Double threat patients had an increased emergency department visit probability with odds ratios of 4.57 (95% CI 4.56-4.58) compared to those not using opioids.

In November 2020, The Agency for Healthcare Research and Quality (AHRQ) released a technical brief on Prevention, Diagnosis, and Management of Opioids, Opioid Misuses, and Opioid Use Disorder in Older Adults. (10) AHRQ identified seven empirical studies between 2000 and 2020 examining the relationship between benzodiazepine use and long-term opioid use. Studies were mostly consistent (≥75% agreement) that concomitant benzodiazepine use was associated with long-term opioid use (3 studies found a strong association, 3 studies found a weak association, and 1 study found no statistically significant association). AHRQ notes that long-term opioid use, though not a harm per se, may increase the risk of harms if not appropriately managed.

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

A primary search of the literature was conducted via PubMed for clinical trials and observational studies (April 2015 through February 2018), and a search of the FDA website was conducted.

The primary search of the literature review described above was completed, widening the search timeframe from 2015 to 2021.

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

Sun EC, Dixit A, Humphreys K, et al. Association between concurrent use of prescription opioids and benzodiazepines and overdose: retrospective analysis. BMJ. 2017;356:j760. doi: 10.1136/bmj.j760. PMID: 28292769

Gaither JR, Goulet JL, Becker WC, et al. The Association Between Receipt of Guideline-Concordant Long-Term Opioid Therapy and All-Cause Mortality. J Gen Intern Med 2016; 31:492

Dasgupta N, Funk MJ, Proescholdbell S, et al. Cohort Study of the Impact of High-Dose Opioid Analgesics on Overdose Mortality. Pain Med 2016; 17:85.

US Food and Drug Administration. FDA Drug Safety Communication: FDA warns about serious risks and death when combining opioid pain or cough medicines with benzodiazepines; requires its strongest warning. August 31, 2016. Available at: http://www.fda.gov/Drugs/DrugSafety/ucm518473.htm. Accessed: November 9, 2016.

Park TW, Saitz R, Ganoczy D, et al. Benzodiazepine prescribing patterns and deaths from drug overdose among US veterans receiving opioid analgesics: case-cohort study. BMJ. 2015; 350:h2698. PMID: 26063215.

Garg RK, Fulton-Kehoe D, Franklin GM. Patterns of Opioid Use and Risk of Opioid Overdose Death Among Medicaid Patients. Med Care. 2017 Jul;55(7):661-668. PMID: 28614178.

Hernandez I, He M, Brooks MM, Zhang Y. Exposure-Response Association Between Concurrent Opioid and Benzodiazepine Use and Risk of Opioid-Related Overdose in Medicare Part D Beneficiaries. JAMA Netw Open. 2018 Jun 1;1(2):e180919. PMID: 30646080.

Lo-Ciganic J, et al. 2018. Geographic Variation of High-Risk Opioid Use and Risk of Overdose Among Disabled Medicare Beneficiaries in the US from 2011 to 2015. Value in Health. 21(2018): S2.

Watanabe JH, Yang J. Association of combination opioid, benzodiazepine, and muscle relaxant usage with emergency department visits in a nationwide cohort in the United States. Int J Clin Pharm. 2020 Apr 7. doi: 10.1007/s11096-020-01012-5. Epub ahead of print. PMID: 32266557.

Zullo AR, Danko KJ, Moyo P, Adam GP, Riester M, Kimmel HJ, Panagiotou OA, Beaudoin FL, Carr D, Balk EM. Prevention, Diagnosis, and Management of Opioids, Opioid Misuse, and Opioid Use Disorder in Older Adults [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2020 Nov. Report No.: 21-EHC005. PMID: 33211447.

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.*

Overdose deaths involving prescription opioids were five times higher in 2016 than in 1999, and more than 200,000 people have died in the U.S. from overdoses related to prescription opioids.(1,2) Scientific research has identified high-risk prescribing practices that have contributed to the opioid overdose epidemic, including overlapping opioid and benzodiazepine prescriptions.(3) Concurrent use of opioids and benzodiazepines, both central nervous system (CNS) depressants, increases the risk for severe respiratory depression, which can be fatal.(3,4)

According to the Centers for Disease Control and Prevention (CDC) Guideline for Prescribing Opioids for Chronic Pain – United States, 2016, clinicians should avoid concurrent prescribing of opioids and benzodiazepines whenever possible.(3) This is a Category A recommendation (applies to all persons; most patients should receive the recommended course of action) and is based on Type 3 evidence (observational studies or randomized clinical trials with notable limitations). In August 2016, the US Food and Drug Administration added concurrent use of opioids and benzodiazepines as a black box warning to prescription opioids (analgesic and cough medicine) and benzodiazepines.(4)

Several studies indicate that concurrent use of opioids and benzodiazepines puts patients at greater risk for a fatal overdose. Three studies of opioid overdose deaths found evidence of concurrent benzodiazepine use in 31%–61% of cases.(5-7) In the United States, the number of opioid overdose deaths involving benzodiazepines increased 14% on average for each year from 2006 through 2011. However, the number of opioid overdose deaths not involving benzodiazepines did not change significantly.(8) A case-cohort study found that concurrent use of benzodiazepines among US veterans raised the risk of drug overdose deaths four-fold (hazard ratio, 3.86, 95% confidence interval [CI] 3.49-4.26) compared with patients not using benzodiazepines.(9) In a large sample of privately insured patients from 2001-2013, opioid users who also used benzodiazepines were at substantially higher risk of an emergency department (ED) visit or hospital admission for opioid overdose (adjusted odds ratio 2.14; 95% CI, 2.05-2.24). If this association is causal, elimination of the concurrent use could reduce the population risk of an ED visit or hospitalization for opioid overdose by 15%.(10)

Despite the risks, concurrent prescriptions for opioids and benzodiazepines are common and increasing. From 2001-2013, concurrent prescribing (overlap of at least one day) increased by nearly 80% (from 9% to 17%) among privately insured patients.(10) In one study, approximately half of the patients received both opioid and benzodiazepine prescriptions from the same prescriber on the same day.(11) In a 2015 analysis of Medicare Part D non-cancer and/or non-hospice patients on opioid therapy, the prevalence of benzodiazepine concurrent use was 24%.(12)

The PQA Concurrent Use of Opioids and Benzodiazepines measure evaluates a process that correlates with increased risk of opioid overdose. Efforts to prevent opioid overdose deaths should include a multi-faceted approach, including strategies that focus on monitoring and reducing opioid prescribing that has an unfavorable balance of benefit and harm for most patient populations. The measure excludes patients with cancer and those in hospice due to the unique therapeutic goals, ethical considerations, increased opportunities for medical supervision, and balance of risks and benefits with opioid therapy.(3)

1. Hedegaard H, Warner M, Miniño AM. Drug overdose deaths in the United States, 1999–2016. NCHS Data Brief, no 294. Hyattsville, MD: National Center for Health Statistics. 2017/ CDC. Wide-ranging online data for epidemiologic research (WONDER). Atlanta, GA: CDC, National Center for Health Statistics; 2016. Available at http://wonder.cdc.gov

2. Frenk SM, Porter KS, Paulozzi LJ. Prescription opioid analgesic use among adults: United States, 1999–2012. NCHS data brief, no 189. Hyattsville, MD: National Center for Health Statistics. 2015.

3. Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016;65(1):1-49. doi:10.15585/mmwr.rr6501e1.

4. US Food and Drug Administration. FDA Drug Safety Communication: FDA warns about serious risks and death when combining opioid pain or cough medicines with benzodiazepines; requires its strongest warning. August 31, 2016. Available at: http://www.fda.gov/Drugs/DrugSafety/ucm518473.htm. Accessed: November 9, 2016.

5. Gomes T, Mamdani MM, Dhalla I a, Paterson JM, Juurlink DN. Opioid dose and drug-related mortality in patients with nonmalignant pain. Arch Intern Med. 2011;171(7):686-691. doi:10.1001/archinternmed.2011.117.

6. Dasgupta N, Funk MJ, Proescholdbell S, Hirsch A, Ribisl KM, Marshall S. Cohort Study of the Impact of High-dose Opioid Analgesics on Overdose Mortality. Pain Med. September 2015. doi:10.1111/pme.12907.

7. Jones CM, McAninch JK. Emergency Department Visits and Overdose Deaths from Combined Use of Opioids and Benzodiazepines. Am J Prev Med. 2015;49(4):493-501. doi:10.1016/j.amepre.2015.03.040.

8. Chen LH, Hedegaard H, Warner M. Drug-poisoning Deaths Involving Opioid Analgesics: United States, 1999-2011. NCHS Data Brief. 2014;(166):1-8.

9. Park TW, Saitz R, Ganoczy D, Ilgen MA, Bohnert ASB. Benzodiazepine prescribing patterns and deaths from drug overdose among US veterans receiving opioid analgesics?: case-cohort study. :1-8. doi:10.1136/bmj.h2698.

10. Sun EC, Dixit A, Humphreys K, et al. Association between concurrent use of prescription opioids and benzodiazepines and overdose: retrospective analysis. BMJ. 2017;356:j760. doi: 10.1136/bmj.j760. PMID: 28292769

11. Hwang CS, Kang EM, Kornegay CJ, Staffa JA, Jones CM, McAninch JK. Trends in the Concomitant Prescribing of Opioids and Benzodiazepines, 2002-2014. Am J Prev Med. 2016:1-10. doi:10.1016/j.amepre.2016.02.014.

12. CMS. Concurrent Use of Opioids and Benzodiazepines in a Medicare Part D Population. May 12, 2016. 2016. https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Concurrent-Use-of-Opioids-and-Benzodiazepines-in-a-Medicare-Part-D-Population-CY-2015.pdf. Accessed December 6, 2016.

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.*

The measure was tested in two different health plan data sources – the Medicare and the Medicaid populations.

For the Medicare population, data used for testing came from the Medicare 5% national sample using data from January 1, 2015 to December 31, 2015. The analysis included 710 Medicare Advantage Prescription Drug plans (MA-PD) and 73 standalone Prescription Drug Plans (PDPs) covering 2,952,360 individuals aged 18 and older.

The Medicare rates ranged from 2.1% (minimum) to 44.7% (maximum). The mean rate was 22.2% with a standard deviation of 7.3%. The 25th percentile was 17.4%, the 50th percentile (median) was 21.4% and the 75th percentile was 27.3%. The interquartile range was 9.9%.

For the Medicaid population, the majority of testing data came from the National Medicaid Analytic eXtract (MAX) data. The data included 322 health plans from 17 states covering 11,745,722 individuals aged 18 and older. In addition, one state Medicaid program with three state-based health plans covering 222,896 individuals 18 years and older was included in the testing using the state's Medicaid administrative claims database.

The Medicaid rates for the national (MAX) data ranged from 0.0% (minimum) to 17.3% (maximum). The mean was 5.0% with a standard deviation of 3.5%. The 25th percentile was 2.4%, the 50th percentile (median) was 4.5% and the 75th percentile was 6.9%. The interquartile range was 4.5%.

For the one state Medicaid program with the three health plans, the rate ranged from 2.8% to 6.3%.

---UPDATED FOR MAINTENANCE---

Performance scores on the measure as specified are provided from implementations in the Medicare Part D Patient Safety Reports, Medicare Part D Display page, and the Medicaid Adult Core Set.

Medicare Part D Patient Safety Reports

The Medicare Part D Patient Safety Reports are made available to all Part D plan sponsors on a monthly basis for the purposes of quality improvement, allowing plan sponsors to monitor their performance on key quality metrics and compare their performance to overall averages. As noted in the testing form, the Medicare Part D Patient Safety Reports represent performance by plan sponsors spanning the full Medicare Part D program. In 2018, per the Chronic Conditions warehouse, approximately 46.7 million patients were enrolled in the Medicare Part D program. Data are provided for the full 2018 data year and partial data from the 2019 data year reflecting the most recent made available to PQA at the time of submission (final year-to-date rates are

generated at the end of July of each year following the final prescription drug event (PDE) submission at the end of June, with the most recent 2019 data available generated from the April 2019 report, year-to-date). Data are provided stratified by line of business (Medicare Advantage Prescription Drug Plan [MAPD], stand-alone Prescription Drug Plan [PDP]), inclusive of contracts with greater than 30 patients in the denominator.

2018 Data (MAPD n=605)

Mean: 19.44%

St. Dev: 6.72%

Percentiles:

100% Max: 50.27%

99%: 40.64%

95%: 31.48%

90%: 27.05%

75% Q3: 22.91%

50% Median: 19.05%

25% Q1: 15.01%

10%: 11.84%

5%: 9.22%

1%: 6.23%

0% Min: 0.79%

2018 Data (PDP n=58)

Mean: 19.36%

St. Dev: 4.78%

Percentiles:

100% Max: 32.99%

99%: 32.99%

95%: 26.59%

90%: 24.78%

75% Q3: 22.86%

50% Median: 20.05%

25% Q1: 15.65%

10%: 13.30%

5%: 11.76%

1%: 8.76%

0% Min: 8.76%

2019 Data (MAPD n=618)

Mean: 17.39%

St. Dev: 6.15%

Percentiles:

100% Max: 49.53%

99%: 35.88%

95%: 29.28%

90%: 24.67%

75% Q3: 20.74%

50% Median: 17.08%

25% Q1: 13.15%

10%: 10.16%

5%: 8.44%

1%: 5.77%

0% Min: 2.44%

2019 Data (PDP n=57)

Mean: 17.44%

St. Dev: 3.98%

Percentiles:

100% Max: 27.56%

99%: 27.56%

95%: 23.18%

90%: 22.00%

75% Q3: 19.97%

50% Median: 17.88%

25% Q1: 14.83%

10%: 11.72%

5%: 10.53%

1%: 8.13%

0% Min: 8.13%

Medicare Part D Display Page

The COB measure was implemented as a Part D Display measure in 2021 (using 2019 data). The Medicare Part D Display page is a public reporting program, which includes measures that have been transitioned from the Star Ratings, new measures that are tested before inclusion into the Star Ratings, or measures displayed for informational purposes only. Display measures are available to the public via the Part C and D Performance Data [https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData] page. Measure rates are reported for all plan sponsors in the Part D program with sufficient data meeting the minimum denominator requirement of 30, except for contracts whose measurement period is prior to one year past the contract's effective date, in which case the contract is marked as "Plan too new to be measured", or contract that were otherwise not required to report the measure to CMS. Performance distributions are provided below.

Please note that the Display data contain information for primarily the same entities over the same course of time as the 2019 Patient Safety data reported above, with a few minor differences. These differences include, as described above, that the Display data do not include contracts flagged as "Plan too new to be measured" or contracts that were otherwise not required to report the measure to CMS. Additionally, the Patient Safety data

do not contain data from the full calendar year. For these reasons, performance distributions are expected to be similar, but not identical.

2019 data (MAPD N=479)

Mean: 17.20%

St. Dev: 5.68%

Percentiles:

100% Max: 39.70%

99%: 35.88%

95%: 28.06%

90%: 23.99%

75% Q3: 20.20%

50% Median: 16.90%

25% Q1: 13.27%

10%: 10.54%

5%: 9.19%

1%: 5.85%

0% Min: 5.01%

2019 Data (PDP n=57)

Mean: 17.43%

St. Dev: 3.98%

Percentiles:

100% Max: 27.61%

99%: 27.61%

95%: 23.38%

90%: 21.99%

75% Q3: 19.97%

50% Median: 17.88%

25% Q1: 14.83%

10%: 11.73%

5%: 10.53%

1%: 8.13%

0% Min: 8.13%

Medicaid Adult Core Set

Although the COB measure is not yet publicly reported in the Medicaid Adult Core Set program, preliminary deidentified data were obtained to support this NQF submission. These data include performance rates from 19 state Medicaid programs that reported on the measure for calendar year 2018, and 1 state that reported data from federal fiscal year 2018. Of these 20 states, 12 provided data on the Medicaid population only, 3 provided data on Medicaid and Dual Eligible populations, 4 provided data on Medicaid and CHIP populations, and 1 provided data on Medicaid, Dual Eligible, and CHIP population. The total measure denominator population across these 20 states was 765,514. Please note that these data did not undergo the program's final quality assurance and review associated with public reporting. Performance distributions are provided below.

2018 data (N=20)

Mean: 19.15%

St. Dev: 5.36%

Percentiles:

100% Max: 32.80%

99%: 32.80%

95%: 29.45%

90%: 25.20%

75% Q3: 22.35%

50% Median: 18.80%

25% Q1: 14.70%

10%: 12.75%

5%: 11.75%

1%: 11.50%

0% Min: 11.50%

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.

---UPDATED FOR MAINTENANCE---

In addition to performance data above, our literature search returned additional information on the prevalence of concurrent use of opioids and benzodiazepines and opportunity for improvement.

A study using 2013-2014 Medicare Part D claims examined geographic variation in the concurrent use of opioids and benzodiazepines among noncancer beneficiaries who used opioids (N=268,678), examining state, hospital-referral region (HRR), and county-level variation. The adjusted probability of concurrent use ranged from 16.7% to 29.6% across states, 12.1% to 37.0% across HRRs, and 0% to 65.2% across counties. (1) Notably, the authors found that state-level variation masks substantial county-level variation: only 18% of counties located in the lowest state quintile were in the lowest county quintile; only 23% of counties located within the highest state quintile were in the highest county quintile. Based on these results, authors call for policies to better understand and monitor concurrent use at the local level.

A study on opioid and benzodiazepine prescribing in 9 states using the 2015 Prescription Behavioral Surveillance System examined deidentified prescription drug management (PDMP) data (N=19,977,642). The study found that 21.6% of patients prescribed an opioid were also prescribed a benzodiazepine, of which 54.9% had concurrent prescriptions (defined as overlapping for at least 7 consecutive days). (2)

Additionally, a cohort study of administrative data (N=4,897,464) in the Medicare Advantage (MA) and commercial populations from 2014-2018 assessed whether the 2016 CDC Guideline for Prescribing Opioids for Chronic Pain was associated with changes in the rate of co-prescription of opioids and benzodiazepines. (3) The study found that among long-term opioid users, co-prescribing (measured as the adjusted percent of person-months with any overlapping days of opioids and benzodiazepines) appears to have modestly declined after the release of the guidelines, with a slope of -0.95 percentage points per year (95% confidence interval [CI] -1.44 to -0.46) in the MA population and -1.06 percentage points per year in the commercial population (95% CI -1.49 to -0.63). However, co-prescribing rates among long-term opioid users at the end of the study left significant room for improvement, with rates of 24.53 (95% CI 23.98-25.09) in the MA population and 22.18 (95% CI 21.49-22.86) in the commercial population.

1) Hernandez I, He M, Zhang Y. Comparing state, regional, and local variation in concurrent opioid and benzodiazepine use. Drug Alcohol Depend. 2018 Oct 1;191:141-144. Epub 2018 Aug 7. PMID: 30099175.

2) Guy GP Jr, Zhang K, Halpin J, Sargent W. An Examination of Concurrent Opioid and Benzodiazepine Prescribing in 9 States, 2015. Am J Prev Med. 2019;57(5):629-636. PMID: 31564606.

3) Jeffery MM, Hooten WM, Jena AB, Ross JS, Shah ND, Karaca-Mandic P. Rates of Physician Co-prescribing of Opioids and Benzodiazepines After the Release of the Centers for Disease Control and Prevention Guidelines in 2016. JAMA Netw Open. 2019 Aug 2;2(8):e198325. PMID: 31373650.

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.*

Disparities data are available for the Medicare population. The testing for the Medicare population came from the Medicare 5% national sample using data from January 1, 2015 to December 31, 2015. The analysis included 710 Medicare Advantage Prescription Drug plans (MA-PD) and 73 standalone Prescription Drug Plans (PDPs) covering 2,952,360 individuals aged 18 and older.

The beneficiary level Low-Income Subsidy (LIS) variable was used to determine disparities in rates for populations with different sociodemographic status. The LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources. The measure rate for the LIS group was 29.9% while the rate for the non-LIS population was significantly lower, at 19.9%.

---UPDATED FOR MAINTENANCE---

Available performance data from measure implementation in the Medicare Part D Patient Safety Reports program, Medicare Part D Display Page, and Medicaid Adult Core Set did not include stratification by population groups.

Disparities data available from original measure testing for the Medicare population, as described above and in the testing form, are provided below. The measure rate for the LIS group was higher than the rate for the non-LIS group. Additionally, rates by age group trended down as age increased, and rates were higher among females than males.

Measure rates by LIS status (Medicare):

LIS: 29.9%

Non-LIS: 19.9%

Measure rates by age band (Medicare):

Ages 18-50: 37.2%

Ages 51-64: 33.8%

Ages 65-84: 19.4%

Ages 85+: 16.7%

Measure rates by sex (Medicare):

Male: 21.3%

Female: 26.5%

Additionally, for the original Medicaid testing described in the testing form, data for age bands and by sex are available and provided below. Rates were highest in the 51-64 age band. Please note that the 65+ is a very

small proportion of the overall sample; please refer to the original testing information (testing form 1.6) for more information.

Measure rates by age band (Medicaid):

Ages 18-50: 3.1%

Ages 51-64: 6.4%

Ages 65-84: 1.3%

Ages 85+: 0.6%

Measure rates by sex (Medicaid):

Male: 3.5%

Female: 3.7%

As more detailed data on these subpopulations become available in performance data from implementations, PQA will further explore the existence of disparities in measure performance. For additional information on disparities, please refer to section 1b.5.

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

---UPDATED FOR MAINTENANCE---

Data from the literature are consistent with potential performance disparities found in testing, with evidence for additional potential disparities by race, disability, and dual-eligibility, among others. However, the literature does not appear to provide substantial evidence for a performance disparity by sex.

In 2016, CMS published an analysis of concurrent use of opioids and benzodiazepines in the Medicare Part D population. Results were stratified by available variables, including line of business, LIS status, age, and disability. (1) Only small differences in concurrent opioid and benzodiazepine rates were observed between those enrolled in PDPs (25.1%) and MA-PDs (22.7%). The rate of concurrent opioid and benzodiazepine use was over 50% higher for LIS opioid users compared to Non-LIS [(30.6% - 19.7%)/19.7% = 55.3%]. A much greater difference in opioid and benzodiazepine concurrent use was observed in those under age 65 years (36.4%) compared to the older age groups (18.7% to 20.0%). The opioid and benzodiazepine concurrent rate among the disabled (i.e., current Medicare enrollment reason is disabled and disabled with end-stage renal disease [ESRD]) was almost double that of the nondisabled (i.e., current Medicare enrollment reason is aged or ESRD alone) opioid users (36.6% vs 19.5%). These findings (with the exception of disabled status which was not available for our measure testing, and sex, which was not examined in the study) are in alignment with data from our measure testing provided in section 1b.4.

A study of utilization patterns for concurrent use of opioids and benzodiazepines among community-dwelling adults from 2011-2015 found that White race was significantly associated with reporting concurrent opioid and benzodiazepine use [odds ratio [OR]=1.65; 95% CI 1.18-2.30]. (2) Additional significant individual-level factors included living in the Southern region of the United States [OR=1.71; 95% CI 1.11-2.63], having a disability [OR=2.68; 95% CI 1.94-3.70], smoking [OR=1.71; 95% CI 1.29-2.26], and several others. Female sex was not found to be a significant factor [OR=.99; 95% CI .75-1.29].

A study of emergency department visits between 2004 and 2011 assessed trends due to nonmedical concurrent use of opioids and benzodiazepines. (3) Throughout the study, there were higher rates of emergency department use among those who were White versus Black or Hispanic, with data from the final study year: White (45.1 visits per 100,000) compared to those whose race was Black (21.0 visits per 100,000) or Hispanic (5.1 visits per 100,000). Rates were not consistently higher across the study period for male versus female sex.

A cohort study of administrative data (n=4,897,464) in the Medicare Advantage and commercial populations from 2014-2018 assessed whether the 2016 CDC Guideline for Prescribing Opioids for Chronic Pain was associated with changes in the rate of co-prescription of opioids and benzodiazepines (4). In secondary analyses, the study found that White and Hispanic beneficiaries had higher rates of co-prescription than Black and Asian beneficiaries over the entire study period. Among long-term opioid use episodes, the adjusted percent of person-months with any overlapping days of opioids and benzodiazepines by race were at the end of the study were: White (24.71; 95% CI 24.19 to 25.22), Hispanic (23.11; 95% CI 21.56-24.67), Black (18.12; 95% CI 17.12-19.11), Asian (18.46; 95% CI 14.35-22.57). Authors did not note differences in rates of overlapping opioids and benzodiazepines by sex.

A retrospective cohort study of Medicare Part D claims data from 2013-2014 examined risk of opioid-related overdose in beneficiaries who filled at least one prescription for an opioid. (5) The study sought to evaluate the exposure-response association between days with concurrent use and risk of overdose. The study found that White, disabled, dual-eligible, and LIS individuals were more likely to use prescription opioids and benzodiazepines concurrently, although relative percentages and hazard ratios were not provided. The authors did not note that male or female sex was associated with a greater likelihood to use prescription opioids and benzodiazepines concurrently.

A cross-sectional analysis of veterans dually-enrolled in the Veterans Health Administration (VA) drug benefit and the Medicare Part D program compared rates of concurrent use of benzodiazepines between those receiving prescriptions exclusively through VA and those receiving prescriptions through VA and the Medicare Part D program in 2013. (6) The outcome measure was the PQA Concurrent Use of Opioids and Benzodiazepines measure; consistent with measure specifications, the study population was patients who received at least two opioid prescriptions on at least two different dates of service with at least fifteen cumulative days' supply during the year. Concurrent use was more frequent in the dual use group versus the VA only (23.1% vs. 17.3%, adjusted risk ratio=1.27; 95% CI 1.24-1.30) and versus Part D only (23.1% vs 16.5%, adjusted risk ratio 1.12; 95% CI 1.10-1.14).

White race as a potential disparity for receiving concurrent use of opioids and benzodiazepines is consistent with a higher rate of opioid overdoses among those whose race was white (26.7 per 100,000), versus those whose race was Black (14.4 per 100,000) or Hispanic (10.0 per 100,000), per a CDC Morbidity and Mortality Weekly Report examining data from 2015-2017. (7)

The studies listed above did not support the existence of a disparity across male versus female sex.

---

1) CMS. Concurrent Use of Opioids and Benzodiazepines in a Medicare Part D Population [Internet]. 2016 [cited 2016 Dec 6]. Available from: https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Concurrent-Use-of-Opioids-and-Benzodiazepines-in-a-Medicare-Part-D-Population-CY-2015.pdf.

2) Vadiei N, Bhattacharjee S. Concurrent Opioid and Benzodiazepine Utilization Patterns and Predictors Among Community-Dwelling Adults in the United States. Psychiatr Serv. 2020 Oct 1;71(10):1011-1019. doi: 10.1176/appi.ps.201900446. Epub 2020 Jun 10. PMID: 32517642.

3) Jones CM, McAninch JK. Emergency Department Visits and Overdose Deaths from Combined Use of Opioids and Benzodiazepines. Am J Prev Med. 2015 Oct;49(4):493-501. doi: 10.1016/j.amepre.2015.03.040. Epub 2015 Jul 3. PMID: 26143953.

4) Jeffery MM, Hooten WM, Jena AB, Ross JS, Shah ND, Karaca-Mandic P. Rates of Physician Co-prescribing of Opioids and Benzodiazepines After the Release of the Centers for Disease Control and Prevention Guidelines in 2016. JAMA Netw Open. 2019 Aug 2;2(8):e198325. PMID: 31373650.

5) Hernandez I, He M, Brooks MM, Zhang Y. Exposure-Response Association Between Concurrent Opioid and Benzodiazepine Use and Risk of Opioid-Related Overdose in Medicare Part D Beneficiaries. JAMA Netw Open. 2018 Jun 1;1(2):e180919. PMID: 30646080.

6) Carico R, Zhao X, Thorpe CT, Thorpe JM, Sileanu FE, Cashy JP, Hale JA, Mor MK, Radomski TR, Hausmann LRM, Donohue JM, Suda KJ, Stroupe K, Hanlon JT, Good CB, Fine MJ, Gellad WF. Receipt of Overlapping Opioid and Benzodiazepine Prescriptions Among Veterans Dually Enrolled in Medicare Part D and the Department of Veterans Affairs: A Cross-sectional Study. Ann Intern Med. 2018 Nov 6;169(9):593-601. doi: 10.7326/M18-0852. Epub 2018 Oct 9. PMID: 30304353; PMCID: PMC6219924.

7) Lippold KM, Jones CM, Olsen EO, Giroir BP. Racial/Ethnic and Age Group Differences in Opioid and Synthetic Opioid-Involved Overdose Deaths Among Adults Aged =18 Years in Metropolitan Areas - United States, 2015-2017. MMWR Morb Mortal Wkly Rep. 2019 Nov 1;68(43):967-973. doi: 10.15585/mmwr.mm6843a3. PMID: 31671083; PMCID: PMC6822810.

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.)*

https://www.pqaalliance.org/measures-overview#cob

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

This is not an eMeasure **Attachment:**

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

Attachment **:** pqa_meas_yr_2019_cob_value_sets_20200729_NQF.xlsx

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

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

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

Not an instrument-based measure

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

Yes

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

As previously noted during the annual update, these specifications have been updated to include a new denominator exclusion for individuals with a diagnosis of sickle cell disease. This change is based on feedback received from measure users, expert input, review and recommendations from PQA's Measure Update Panel, and PQA's Quality Metrics Expert Panel's approval of the exclusion recommendation. Individuals with sickle cell disease have unique pain management needs, and the Centers for Disease Control have stated that their Guideline for Prescribing Opioids for Chronic Pain is not intended to apply to patients with sickle cell disease [Available at https://www.asco.org/sites/new-www.asco.org/files/content-files/advocacy-and-policy/documents/2019-CDC-Opioid-Guideline-Clarification-Letter-to-ASCO-ASH-NCCN.pdf]. Due to these considerations, and their unique therapeutic goals, ethical considerations, opportunities for medical supervision, and balance of risks and benefits, individuals with a diagnosis of sickle cell disease are excluded from this measure.

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 number of individuals from the denominator with concurrent use of opioids and benzodiazepines for >=30 cumulative days during the measurement year.

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 number of individuals from the denominator with:

- >=2 prescription claims for any benzodiazepine with different dates of service, AND
- Concurrent use of opioids and benzodiazepines for >=30 cumulative days.

Complete the steps below to identify individuals with concurrent use of opioids and benzodiazepines:

Step 1: From the denominator population, identify individuals with >=2 prescription claims with different dates of service for any benzodiazepine (Table COB-B, below) during the measurement year.

Step 2: Determine the total days of overlap (concurrent use) between the opioid and benzodiazepine prescription claims during the measurement year.

Concurrent use is identified using the dates of service and days' supply of an individual's opioid and benzodiazepine prescription claims. The days of concurrent use is the count of days during the measurement year with overlapping days' supply for an opioid and a benzodiazepine. Exclude days' supply and overlap that occur after the end of the measurement year.

NOTE:

If multiple prescriptions for opioids (or benzodiazepines) are dispensed on the same day, calculate the number of days covered by an opioid (or benzodiazepine) using the prescription claims with the longest days' supply.

If multiple prescription claims of opioids (or benzodiazepines) are dispensed on different days with overlapping days' supply, count each day in the measurement year only once toward the numerator. There is no adjustment for early fills or overlapping days' supply for opioids (or benzodiazepines).

Step 3: Count the individuals with concurrent use for >=30 cumulative days. This is the numerator.

Table COB-B: Benzodiazepines:

Alprazolam, chlordiazepoxide, clobazam, clonazepam, clorazepate, diazepam, estazolam, flurazepam, lorazepam, midazolam, oxazepam, quazepam, temazepam, triazolam

(Note: excludes injectable formulations, includes combination products)

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

The denominator includes individuals >=18 years of age with >=2 prescription claims for opioid medications on different dates of service and with >=15 cumulative days' supply during the measurement year. Individuals with cancer, sickle cell disease, or in hospice are excluded.

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 individuals 18 years and older by the first day of the measurement year with >=2 prescription claims for opioid medications on different dates of service and with >=15 cumulative days' supply during the measurement year. Use Table COB-A: Opioids, below, to identify the opioid medications for the measure.

Complete the steps below to determine the denominator:

Step 1: Identify individuals >=18 years of age as of the first day of the measurement year.

Step 2: Identify individuals meeting the continuous enrollment criteria.

To be continuously enrolled, an individual must be enrolled for the entire measurement year with no more than one gap in enrollment of up to 31 days during the measurement year. When enrollment is verified monthly, the individual may not have more than a 1-month gap in coverage (i.e., an individual whose coverage lapses for 2 months [60 days] is not considered continuously enrolled).

Step 3: Identify individuals with an Index Prescription Service Date (IPSD) that is >=30 days from the last day of the measurement year (January 1 through December 2). The IPSD is defined as the earliest date of service for an opioid during the measurement year.

Step 4: Identify individuals with >=2 prescription claims for opioids on different dates of service, and with >=15 cumulative days' supply during the measurement year. Exclude days' supply that occur after the end of the measurement year.

NOTE:

The prescription can be for the same or different opioids.

If multiple prescriptions for opioids are dispensed on the same day, calculate the number of days covered by an opioid using the prescription claims with the longest days' supply.

If multiple prescriptions for opioids are dispensed on different days, sum the days' supply for all the prescription claims, regardless of overlapping days' supply.

Table COB-A: Opioids:

Benzhydrocodone, buprenorphine, butorphanol, codeine, dihydrocodeine, fentanyl, hydrocodone, hydromorphone, levorphanol, meperidine, methadone, morphine, opium, oxycodone, oxymorphone, pentazocine, tapentadol, tramadol

(note: includes combination products and prescription opioid cough medications. Excludes the following: injectable formulations; sublingual sufentanil (used in a supervised setting); and single-agent and combination

buprenorphine products used to treat opioid use disorder (i.e., buprenorphine sublingual tablets, Probuphine® Implant kit subcutaneous implant, and all buprenorphine/naloxone combination products).

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

Individuals with cancer, sickle cell disease, or in hospice at any point during the measurement year are excluded from the denominator.

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.)*

Hospice exclusion: Exclude any individual in hospice during the measurement year. To identify individuals in hospice:

Use the hospice indicator from the enrollment database, where available (e.g. Medicare); or

>=1 claim, encounter, or medical record during the measurement year. See Hospice Encounter Value Set and Hospice Intervention Value Set (e.g., Medicaid, commercial).

Cancer exclusion: Exclude any individuals with cancer during the measurement year. To identify individuals with cancer:

>=1 claim with cancer in the primary diagnosis or any other diagnosis fields during the measurement year. See Value Set, Cancer.

Sickle Cell Disease exclusion: Exclude any individual with sickle cell disease during the measurement year.

=1 claim with sickle cell disease (SCD) in the primary diagnosis or any other diagnosis fields during the measurement year. See Value Set, Sickle Cell Disease.

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.)*

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.)*

A. Target population (denominator):

Step 1: Identify individuals >=18 years of age as of the first day of the measurement year.

Step 2: Identify individuals meeting the continuous enrollment criteria.

To be continuously enrolled, an individual must be enrolled for the entire measurement year with no more than one gap in enrollment of up to 31 days during the measurement year. When enrollment is verified monthly, the individual may not have more than a 1-month gap in coverage (i.e., an individual whose coverage lapses for 2 months [60 days] is not considered continuously enrolled).

Step 3: Identify individuals with an Index Prescription Service Date (IPSD) that is >=30 days from the last day of the measurement year (January 1 through December 2). The IPSD is defined as the earliest date of service for an opioid during the measurement year.

Step 4: Identify individuals with >=2 prescription claims for opioids on different dates of service, and with >=15 cumulative days' supply during the measurement year. Exclude days' supply that occur after the end of the measurement year.

NOTE:

The prescription can be for the same or different opioids.

If multiple prescriptions for opioids are dispensed on the same day, calculate the number of days covered by an opioid using the prescription claims with the longest days' supply.

If multiple prescriptions for opioids are dispensed on different days, sum the days' supply for all the prescription claims, regardless of overlapping days' supply.

Step 5: Identify individuals with cancer, sickle cell disease or in hospice during the measurement year.

To identify individuals in hospice:

Use the hospice indicator from the enrollment database, where available (e.g., Medicare); or

>=1 claim, encounter, or medical record during the measurement year. See Hospice Encounter Value Set and Hospice Intervention Value Set (e.g., Medicaid, commercial).

To identify individuals with cancer:

>=1 claim with cancer in the primary diagnosis or any other diagnosis fields during the measurement year. See Value Set, Cancer.

To identify individuals with sickle cell disease:

>=1 claim with sickle cell disease (SCD) in the primary diagnosis or any other diagnosis fields during the measurement year. See Value Set, Sickle Cell Disease.

Step 6: Exclude individuals with cancer, sickle cell disease, or in hospice (Step 5) from those identified in Step 4. This is the denominator.

B. Numerator Population:

Step 7: From the denominator population, identify individuals with >=2 prescriptions claims with different dates of service for any benzodiazepines (Table COB-B, below) during the measurement year.

Step 8: Determine the total days of overlap (concurrent use) between the opioid and benzodiazepine prescription claims during the measurement year.

Concurrent use is identified using the dates of service and days' supply of an individual's opioid and benzodiazepine prescription claims. The days of concurrent use is the count of days during the measurement year with overlapping days' supply for an opioid and a benzodiazepine. Exclude days' supply and overlap that occur after the end of the measurement year.

Note: When identifying days' supply for opioids (or benzodiazepines):

If multiple prescriptions for opioids (or benzodiazepines) are dispensed on the same day, calculate the number of days covered by an opioid (or benzodiazepine) using the prescription claims with the longest days' supply.

If multiple prescription claims of opioids (or benzodiazepines) are dispensed on different days with overlapping days' supply, count each day in the measurement year only once toward the numerator. There is no adjustment for early fills or overlapping days' supply for opioids (or benzodiazepines).

Step 9: Count the number of individuals with concurrent use of opioids and benzodiazepines for >=30 cumulative days. This is the numerator.

C. Measure Rate:

Step 10: Divide the number of individuals in the numerator (Step 9) by the denominator (Step 6) and multiply by 100. This is the measure rate reported as a percentage.

Report the rates separately by line of business (e.g. Medicare, Medicaid, Commercial).

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, Enrollment Data

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

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

Administrative claims: prescription claims, medical claims, enrollment data

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

No data collection instrument provided

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

Health Plan

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

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

EmpiricalValidity_COB_OutcomeCodes_FINAL-637401674147076503.xlsx,COB_NQFTestingForm_11062020_FINAL_NQFFEEDBACK.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing

attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

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

No - This measure is not risk-adjusted.

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

Measure Number (*if previously endorsed*)**:** 3389
Measure Title: Concurrent Use of Opioids and Benzodiazepines
Date of Submission: **11/6/2020**

Type of Measure:

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

*cell intentionally left blank

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

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

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

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	Measure Tested with Data From:
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims

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

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

The measure was tested in two different health plan data sources – the Medicare and the Medicaid populations.

For the Medicare population, data used for testing came from the Medicare 5% national sample data. The Medicare Part D Prescription Drug Event (PDE) claims were used for the identification of prescription drugs. The 5% medical claims (standard analytic files) were used to identify cancer diagnoses and hospice claims. To identify dates of birth and continuous enrollment, the Medicare Beneficiaries Summary Files (MBSF) were used.

For the Medicaid population, the data used for testing came from Medicaid administrative claims. National Medicaid sample data covering 31 states and 295 health plans were included in the testing using data from the Medicaid Analytic eXtract (MAX) data. In addition, one state Medicaid program with three state-based health plans was included in the testing using the state's Medicaid administrative claims database. Medical claims were used to identify the cancer diagnoses, and the pharmacy claims were used for the identification of prescription drugs.

Sickle cell disease exclusion testing was completed using data from a representative 2019 sample of a major health plan across the Medicare Advantage Part D (MAPD), Prescription Drug Plan (PDP), and commercial lines of business.

Updated reliability testing was completed based on 2018 performance data in the Medicare Part D Patient Safety Reports.

Empirical validity testing was completed using a 2016 5% Medicare sample.

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

The testing for the Medicare population used administrative claims data from January 1, 2015 to December 31, 2015. The testing for Medicaid used administrative claims data from January 1, 2008 to December 31, 2008 for the national level MAX dataset, and data from January 1, 2016 to December 31, 2016 for one state-based Medicaid dataset. The data from these time periods were the most recent, complete, full year data available to testers at the time of testing.

For sickle cell disease exclusion testing, all data were from the 2019 calendar year.

For the updated reliability testing, the most complete available data for the Medicare Part D Patient Safety Reports are from program year 2018.

For empirical validity testing, data are from the 2016 calendar year.

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

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

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

The Medicare testing was conducted using the Medicare 5% sample data – a nationally representative sample, including data from all the states. Of beneficiaries aged 18 years or older by the first day of the measurement year, the data included 710 Medicare Advantage Prescription Drug (MA-PD) plans and 73 standalone Prescription Drug Plans (PDPs). Plans varied in size (see Table 1), with a mean plan size of 2,639 beneficiaries and a median plan size of 353 beneficiaries.

Table 1. Plan Size Distribution for 2015 Medicare Sample

Statistic	Number of Beneficiaries
Mean	2,639
Standard Deviation	14,308
Minimum	1
25th Percentile	44
50th Percentile	353
75th Percentile	1,264
Maximum	228,698
Interquartile Range	1,220

For the Medicaid testing, the national level analysis included 295 health plans covering 31 states with beneficiaries aged 18 years or older. Of the 295 plans, 31 plans were fee-for-service (FFS), and the remaining

264 plans were Medicaid Managed Care plans. There was a lot of variation in plan size, with mean plan size of 40,372 beneficiaries, and a median plan size of 5,778 beneficiaries (see Table 2). Twenty-four plans were from the Midwest region of the United States (US), 46 plans were from the Northeast region of the US, 111 plans were from states in the South region of the US, and 114 plans were from the West region of the US.

Table 2. Plan Size Distribution for 2008 Medicaid MAX Sample

Statistic	Number of Beneficiaries
Mean	40,372
Standard Deviation	115,401
Minimum	1
25th Percentile	500
50th Percentile	5,778
75th Percentile	25,829
Maximum	1,130,260
Interquartile Range	25,329

The one state-based Medicaid program was in the South region and included 3 health plans – 1 FFS and 2 managed care plans. The mean size of the plans was 74,299 beneficiaries.

For sickle cell disease exclusion testing, data included 57 MAPD contracts, 3 PDP contracts, and 43 commercial contracts.

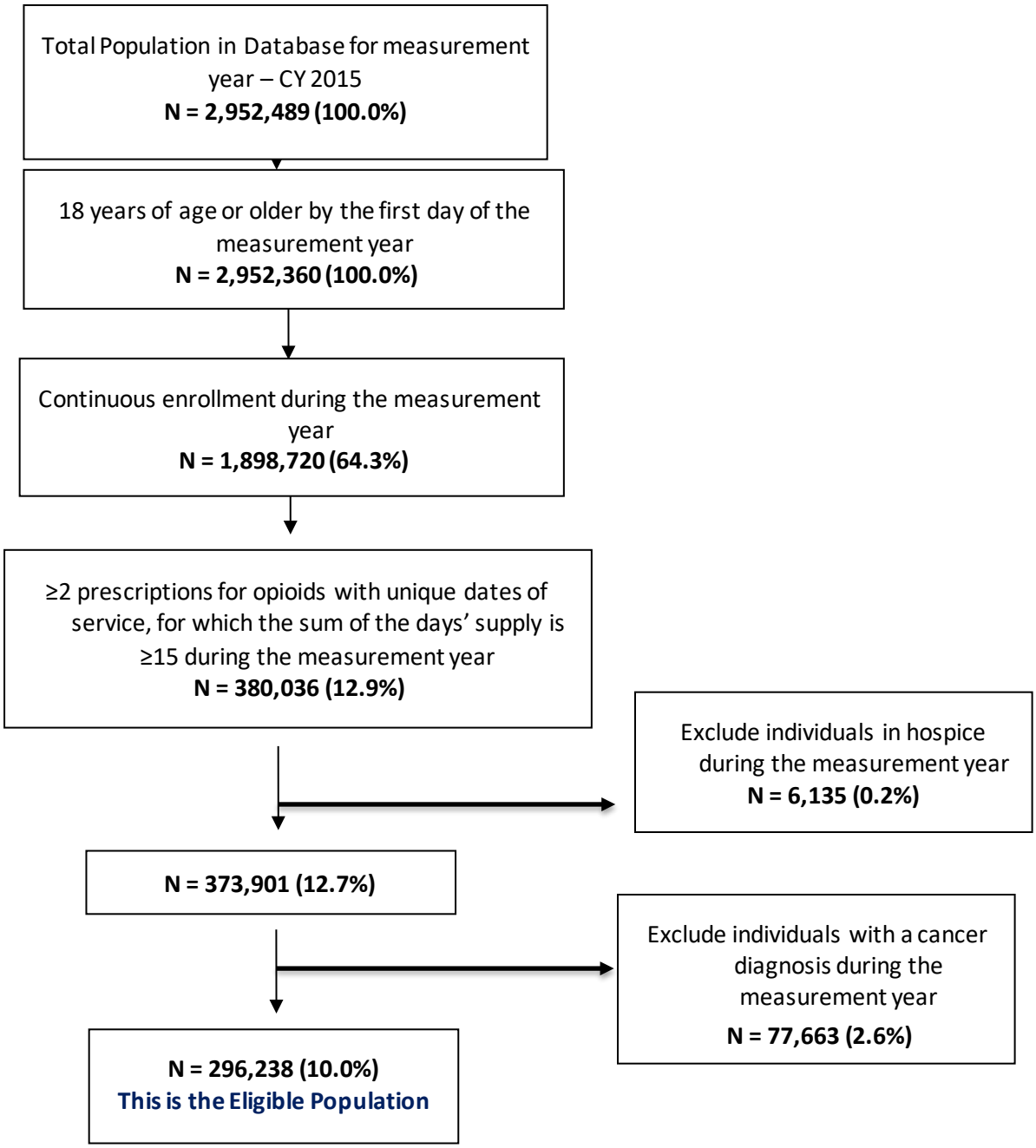
For the updated reliability testing, the Medicare Part D Patient Safety Reports included 676 MAPD contracts and 63 PDP contracts.

For empirical validity testing, the 2016 5% sample included 340 MAPD contracts and 50 PDP contracts.

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

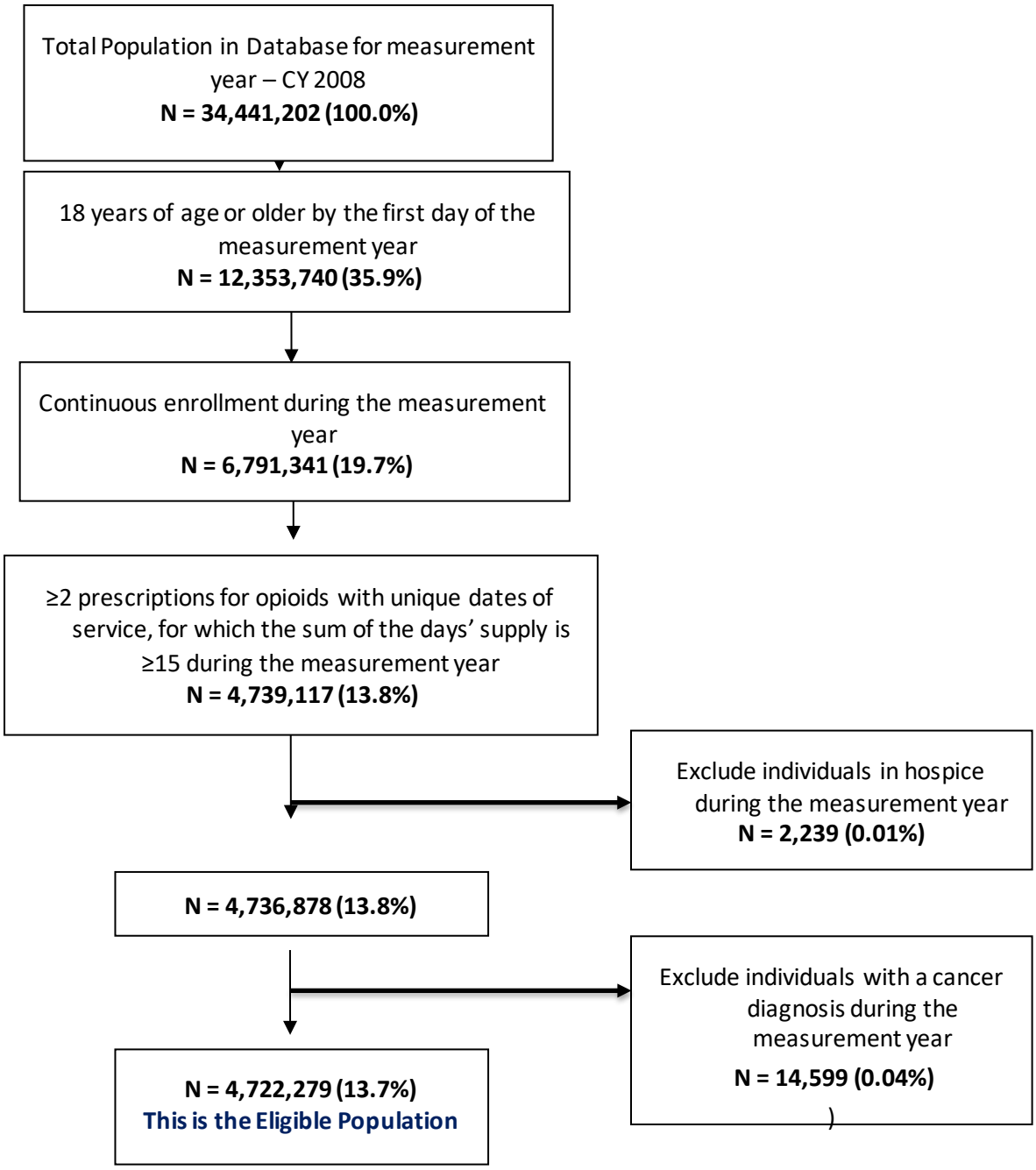
For the Medicare testing, a total of 2,952,360 individuals aged 18 and older were included in the testing and analysis. Of all persons, 1,339,615 (45.4%) were male, and 1,612,745 (54.6%) were female. Individuals by age group included 176,663 (6.0%) age 18 – 50 years, 459,964 (15.6%) age 51 – 64 years, 2,017,849 (68.3%) age 65 – 84 years, and 297,884 (10.1%) age 85 and older. After applying all inclusion and exclusion criteria, the final population for analysis was 296,238 (10.0%) of the initial population. See Figure 1, for the selection criteria for the eligible population for Medicare.

Figure 1. Selection Criteria for Eligible Population for the 2015 Medicare Sample



For the Medicaid MAX population, a total of 12,353,740 beneficiaries aged 18 and older were included in the analysis. Of all persons, 3,489,761 (28.2%) were male, and 8,863,979 (71.8%) were female. Individuals by age group included 10,711,475 (86.7%) age 18 – 50 years, 1,433,050 (11.6%) age 51 – 64 years, 196,151 (1.6%) age 65 – 84 years, and 13,064 (0.1%) age 85 and older. After applying all inclusion and exclusion criteria, the final population for analysis was 4,722,279 (13.7%) of the initial population. See Figure 2, for the selection criteria for the eligible population for the Medicaid MAX population.

Figure 2. Selection Criteria for Eligible Population for 2008 Medicaid MAX Sample



Finally, for the 1 state-based Medicaid program, a total of 222,896 beneficiaries aged 18 and older were included in the analysis. Of all persons, 53,944 (24.2%) were male, and 168,952 (75.8%) were female. Individuals by age group included 183,647 (82.4%) age 18 – 50 years, 36,535 (16.4%) age 51 – 64 years, 2,614 (1.2%) age 65 – 84 years, and 100 (0.04%) age 85 and older. After applying all inclusion and exclusion criteria, the final population for analysis was 99,390 (14.3%) of the initial population.

As seen in the results above, the measure was tested across a large spectrum of age groups, with the Medicare population being older (primarily 65 years and older), and the Medicaid data looking at a much younger population.

For sickle cell disease exclusion testing, a total of 3,952,888 patients were included in the MAPD line of business, a total of 4,854,234 patients were included in the PDP line of business, and a total of 14,270,346 patients were included in the commercial line of business.

The Medicare Part D Patient Safety Reports represent performance by plans spanning the full Medicare Part D program. In 2018, per the Chronic Conditions warehouse, approximately 46.7 million patients were enrolled in the Medicare Part D program. (1)

Medicare Part D Charts. Chronic Conditions Warehouse. N.d. Available from https://www2.ccwdata.org/web/guest/medicare-charts/medicare-part-d-charts.

The 2016 5% Medicare sample used for testing included a total population of 3,039,983 patients across the MAPD and PDP lines of business.

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.

Reliability testing was conducted for both the Medicare and Medicaid populations. For the Medicare population, reliability testing was conducted at the plan contract level, as the application of this measure in the Medicare program would be assessed at the plan contract level. In accordance with the PQA measure specifications, reliability testing excluded plan contracts with less than 30 individuals in the denominator.

For the Medicaid population, reliability testing was conducted at the plan level using the MAX data, and excluded any plans with less than 30 individuals in the denominator.

As noted above, the added sickle cell disease exclusion was tested separately using a representative 2019 sample from a major health plan, updated reliability was tested using the data from the 2018 Part D Patient Safety Reports, and empirical validity was tested using data from the 2016 5% Medicare sample.

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.

For the Medicare population, the beneficiary level Low Income Subsidy (LIS) variable was used to determine disparities in rates for populations with different sociodemographic status. The LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources.

For the Medicaid populations, no patient level indicators of sociodemographic status were available in the data.

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*)

Using the Medicare and Medicaid data described in sections 1.2 to 1.6, the reliability of the computed measure scores was measured as the ratio of signal-to-noise. The signal is the proportion of the variability in measured performance that can be explained by true differences in plan (or contract) performance. Reliability scores range from 0 to 1, with a score of 0 signifying that all variation is due to measurement error. A value of 1 signifies that the variation represents true differences in performance scores between plans. A reliability score of 0.7 is the minimum threshold for reliability.

A beta-binomial model was used to calculate plan-specific reliability scores. This is based on the methods outlined by Adams in the following paper: Adams JL. The reliability of provider profiling: a tutorial. Santa Monica, CA: RAND Corporation. 2009. Retrieved from http://www.rand.org/pubs/technical_reports/TR653.

The reliability score is defined as the ratio of the plan-to-plan variance to the sum of the plan-to-plan variance and the plan-specific error. The plan-to-plan variance is an estimate of the variance of the true rates. The plan-specific error variance is the sampling or measurement error.

$$reliability = \frac{\sigma^2_{plan-to-plan}}{\sigma^2_{plan-to-plan} + \sigma^2_{plan-specific-error}}$$

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*)

Using the parameter estimates from the beta-binomial model, we computed individual plan (or contract) reliability scores. Table 3 shows the distribution of the plan contract-level scores for Medicare, and Table 4 shows the plan-level scores for Medicaid.

Table 3. Plan Contract Reliability Score Distribution for 2015 Medicare Sample

Statistic	Values
Mean	0.7730

Statistic	Values
Standard Deviation	0.1601
Minimum	0.3628
25th Percentile	0.6569
50th Percentile	0.7995
75th Percentile	0.9153
Maximum	0.9986
Interquartile Range	0.2584

The mean reliability score for the Medicare plan-contracts is 0.7730, and the median is 0.7995. Reliability is affected in part by sample size, and as shown for the Medicare contracts distribution in Table 1, the median plan-contract size is 353 beneficiaries.

In contrast, the median plan distribution for the Medicaid population is much larger – 5,778 beneficiaries (see Table 2). Medicaid plans have very high reliability scores. The mean reliability score in the Medicaid plans is 0.9370, and the median is 0.9953 (see Table 4).

Table 4. Plan Reliability Score Distribution for 2008 Medicaid MAX Sample

Statistic	Values
Mean	0.9370
Standard Deviation	0.1871
Minimum	0.0124
25th Percentile	0.9717
50th Percentile	0.9953
75th Percentile	0.9996
Maximum	1.0000
Interquartile Range	0.0279

In order to demonstrate reliability in the COB measure's implementation in the field, PQA conducted reliability analyses on data from the 2018 Part D Patient Safety Reports using the Adams beta-binomial methodology described above. Estimates were only computed for contracts with greater than 30 patients in the denominator. Table 4A provides the distribution of reliability estimates by line of business.

Table 4A. Plan Reliability Score Distribution for the Part D Patient Safety Reports

Statistic	Values (MAPD)	Values (PDP)
10th Percentile	.53	.72
25th Percentile	.79	.89
Median	.95	.98
75th Percentile	.99	.996
90th Percentile	.995	.999
Mean	**.86**	**.91**
Standard Deviation	**.18**	**.15**

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?*)

A reliability score of 0.7 is the minimum threshold for reliability. Based on the mean reliability score of 0.77 for Medicare and 0.94 for Medicaid, the measure is considered reliable.

Based on a mean reliability score of .91 for PDPs and .85 for MAPDs, the measure is considered reliable as used in the Medicare Part D Patient Safety Reports.

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)*

Systematic assessment of face validity

PQA uses a systematic, transparent, consensus-based measure development, testing, and endorsement process. That process used in 2016 to develop this measure is outlined below:

- **Step 1**: Measure concepts for development are prioritized by PQA staff based on input from PQA's Measure Advisement Group, Implementation Advisory Panel, and Patient and Caregiver Advisory Panel. Environmental scans are conducted to identify whether similar measures exist, ensuring harmonization and avoiding duplication. Selected concept ideas are considered to represent areas in which there are measurement and performance gaps to have the greatest chance of implementation in existing measure sets and performance systems, and to align with the National Quality Strategy.

- **Step 2**: PQA Measure Development Teams (MDTs) and Task Forces (TFs), comprised of experts in all phases of drug use and management, discuss and draft specifications for measure concepts that may be appropriate for development into fully specified performance measures. The MDTs/TFs focus on specific aspects of the medication-use system and/or specific therapeutic areas and benefit by having their development work reviewed by larger groups, Stakeholder Advisory Panels. They may also receive input from the Patient & Caregiver Advisory Panel, Implementation Advisory Panel, and Risk Adjustment Advisory Panel.

- **Step 3**: PQA MDTs/TFs recommend measure concepts to the PQA Quality Metrics Expert Panel (QMEP) for evaluation and refinement. The QMEP reviews the measure concepts to provide an initial assessment of the key properties of performance measures (i.e., importance, scientific acceptability, feasibility and usability). The measure concepts that are rated highly on these key properties will undergo testing and possibly further technical specification as draft measures.

- **Step 4**: The draft measures are provided to PQA member organizations for their comments prior to preparing technical specifications (including National Drug Code [NDC] lists) for pilot testing. PQA staff use member comments and MDT/TF and QMEP recommendations to formulate a testing plan for each draft measure.

- **Step 5**: PQA selects partners to test the draft measures. These partners are often PQA member health plans or academic institutions with expertise in quality and performance measure testing that also have access to the data sources needed to calculate the measure rates. The testing partner implements the draft technical specifications within their existing datasets and provides a report to PQA that details testing results and recommendations for modifications of the technical specifications.

- **Step 6**: The QMEP reviews the testing results and recommendations and determines final criteria for the measure based on the findings. The QMEP provides a final assessment of the feasibility and reliability of the draft measures.

- **Step 7**: The Measure Validity Panel, an independent group of individuals not involved in the development or review of the measure concept or draft measure, determines through discussion and vote whether the performance measure score is an accurate reflection of quality and can distinguish good from poor performance (i.e., face validity).

- **Step 8**: Performance measures that are recommended by the QMEP for endorsement consideration by the PQA membership are posted on the PQA web site for member review, written comments are requested, and a webinar for member organizations is held to gather feedback and address any questions. This process allows members to discuss their views on the measures in advance of the voting period.

- **Step 9**: PQA member organizations vote on endorsement of the performance measures.

As part of maintenance of endorsement, PQA completed an empirical assessment of measure validity.

The empirical validity of the measure score was assessed using a criterion validity approach, a methodology that evaluates the extent to which performance on a quality measure is associated with conceptually and clinically related outcomes. Specifically, our assessment evaluated the correlation between plan-level performance on the COB measure as specified, and plan-level rates of a composite of inpatient stays and emergency department utilization due to opioid- and benzodiazepine-related adverse events (OBRAEs). This analysis is based on the expected convergent relationship between measure rates and OBRAEs; the better a given plan performs on the COB measure (i.e. lower rate), the lower plan-level rates of OBRAEs are hypothesized to be.

The composite of OBRAEs was developed by adapting a list of opioid-related adverse event codes ICD9/ICD10 and CPT codes originally published in the literature by Digmann et al (1) for use by a Quality Improvement

Organization. This list was further refined using terms identified by searching the Value Set Authority Center (VSAC) for existing value sets capturing OBRAEs including overdose, poisoning, opioid use disorder, and other related codes such as respiratory depression and syncope. The codes were further expanded upon using CPT codes used in a related study by Zedler et al (2). The final list of codes used in the analysis is attached as an appendix to this maintenance submission, titled EmpiricalValidity_COB_OutcomeCodes_Final.

To calculate the rate of OBRAEs within a plan, only individuals within the eligible population of the COB measure were examined. This ensured that comparisons between the COB measure rate and the OBRAE rates were appropriate. Therefore, the OBRAE rate can be conceptualized as a 'measure' where the denominator is the COB eligible population, and the numerator is the number of individuals from the denominator with at least one inpatient or emergency department (ED) claim with an OBRAE-related code.

Only OBRAE codes present in the principal position on claims were included to ensure that they were the main event driving the inpatient or ED stay. For codes that are often related to opioids or benzodiazepines but may have other causes (e.g. respiratory depression and syncope), an accompanying opioid- or benzodiazepine-specific code in a secondary position was required for the claim to be counted in the analysis. Only inpatient and ED claims were used as the outcomes of interest to narrow the analysis to events of greater severity, versus also including outpatient claims. To align with the COB measure calculation, individuals were counted in the denominator/numerator of the OBRAE composite only once, regardless of whether individuals experienced multiple events.

The correlation between the OBRAE rate and the COB rate was evaluated using Spearman's rank-order correlation coefficient. Spearman's was determined to be the appropriate test of correlation given that it is a non-parametric test and the underlying rates are not to be assumed to be normally distributed.

The data source used for this analysis was a 5% Medicare sample from 2016 including PDPs (N=50 contracts) and MAPDs (N=380 contracts). Correlations were analyzed using OBRAE rates and measure rates from the same measurement year. While additional analyses to gauge predictive validity (e.g. measure rate correlation with OBRAE rates in the subsequent year) may be insightful in the future, data limitations prevented these analyses from being conducted in this submission.

1) Digmann R, et al. Use of Medicare Administrative Claims to Identify a Population at High Risk for Adverse Drug Events and Hospital Use for Quality Improvement. J Manag Care Spec Pharm, 2019 Mar;25(3):402-410.

2) Zedler B, et al. Risk Factors for Serious Prescription Opioid-Related Toxicity or Overdose among Veterans Health Administration Patients. Pain Medicine, 2014 Nov;15(11):1911-1929.

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

The measure was assessed for face validity (i.e., whether it appears to measure what it intends to measure) through review by the team that developed the measure (PQA Measure Development Team [MDT] 13: Concurrent Use of Opioids and Benzodiazepines), the PQA Quality Metrics Expert Panel (QMEP), the Measure Validity Panel (MVP), and PQA's full membership. In addition, feedback about validity of the measure was sought out by the two PQA member organizations who tested the measure using their own data, and four external subject matter experts.

MDT 13 was composed of 27 PQA members. After the MDT completed development of the measure specifications, the group voted to determine if the measure concept should continue with further development and review by the PQA QMEP. Out of 27 members of the MDT who voted, 92.5% recommended that the measure move on for QMEP review.

The PQA QMEP is a panel that includes individuals with expertise and experience in pharmacy, medicine, research, and clinical or other technical expertise related to quality improvement and measure development. The names and credentials of the 21 QMEP members in 2016 are listed in Table 5. The QMEP reviewed the measure prior to testing to ensure the importance and usefulness of the draft measure. Specifically, they confirmed that evidence supported that concurrent prescribing of opioids and benzodiazepines was common

and associated with overdose deaths. The QMEP reviewed the results of the measure testing including the performance measure scores reported by plans referenced in Section 2b4 (below). Out of the 16 members of the QMEP who voted, 93.8% recommended that the draft measure be considered for endorsement by the PQA membership, considering the criteria of importance, scientific acceptability, feasibility, and usefulness.

Table 5. PQA 2016 Quality Metrics Expert Panel (QMEP)

QMEP Member Name	QMEP Member Organization
Amanda Brummel, PharmD	Fairview
Bimal Patel, PharmD	MedImpact
Catherine Coast, PharmD	Highmark
Christopher Dezii, RN, MBA, CPHQ	Bristol-Myers Squibb
Christopher Powers, PharmD	CMS
Craig Schilling, PharmD	Optum/UHG
David Nau, PhD, RPh, CPHQ	PQS
Gary Erwin, PharmD	CVS Health
Jenny Weber, PharmD, MS, PCPS, CGP, BCACP	Humana
Jessica Frank, PharmD	OutcomesMTM
Karen Farris, PhD	University of Michigan
Keith Widmer, RPh, BCPP	Express Scripts
Kent Summers, PhD, RPh	Astellas
Lynn Deguzman, PharmD, CGP	Kaiser Permanente
Mary Ann Kliethermes, PharmD	Midwestern University
Mitzi Wasik, PharmD, BCPS	Coventry Health Care/Aetna
Pat Gleason, PharmD, BCPS	Prime Therapeutics
Steve Riddle, PharmD, BCPS	Wolters Kluwer Health
Steven Burch, PhD, RPh	GlaxoSmithKline
Tony Willoughby, PharmD	HealthMart-McKesson
Tripp Logan, PharmD	MedHere Today

After QMEP approval, the draft measure was reviewed by the MVP. The MVP is made up of an independent group of individuals not involved in the development or review of the measure concept or draft measure. Through discussion and vote, the MVP determines whether the performance measure scores have face validity. Of the 6 MVP members who voted, 100% agreed or strongly agreed that the scores obtained from the measure as specified will provide an accurate reflection of quality, and can be used to distinguish good and poor quality between health plans.

PQA membership was notified in November 2016 of the opportunity to consider and vote on endorsement of the performance measure. (Note: PQA membership is comprised of health plans, community pharmacy, long-term care pharmacies, health information technology companies, pharmacy benefit managers, healthcare quality and standards organizations, professional and trade associations, government agencies, and others.) Members received the measure description, key points and supporting evidence, measure specifications, and

the performance measure scores reported by the plans. Voting options included, "Agree" (indicating that the organization approved endorsement of the measure), "Disagree" (indicating that the organization opposed endorsement of the measure) and "Abstain." Out of the 93 PQA member organizations that cast a vote either in favor of or opposed to endorsement, 89% voted in favor of endorsing the measure.

In addition to this process, 100% of the two PQA member organizations who tested the measure using their own data strongly agreed that the measure reflected the quality of care provided for their population.

The opinion of four subject matter experts was sought in July 2016 for input on the measure elements and assessment of the measure overall. The experts were: Deborah Dowell, MD, MPH, Centers for Disease Control & Prevention; Christopher Jones, PharmD, US Department of Health and Human Services; Joshua Sharenstein, MD, Associate Dean, Johns Hopkins Bloomberg School of Public Health; and Don Teater, MD, Teater Health Solutions (previously, National Safety Council). All four subject matter experts were strongly supportive of the measure.

Within the Medicare 5% sample, the Spearman's correlation coefficient was .45 within PDPs (moderate) [p<.0001] and .21 for MAPDs (weak) [p=.001].

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?*)

Based upon the systematic, consensus-based PQA measure development process designed to assure face validity, the measure has been determined to have face validity.

The correlations produced in our criterion validity analyses suggest a statistically significant and moderate-strength relationship between the measure rate and incidence of opioid- and benzodiazepine-related adverse events. Correlations were stronger for PDPs than MAPDs.

Analyses to correlate measure rates to outcomes, particularly outcomes involving inpatient stays and ED utilization, will always encounter noise. There are numerous factors that may contribute to patients experiencing (or not experiencing) these events, and quality measures such as COB are only able to capture one potential contributor: namely, receiving concurrent days' supply for opioids and benzodiazepines. Given the numerous factors that can contribute to these outcomes, very high correlations would be unexpected. However, the correlations found in this analysis do demonstrate a consistent, statistically significant relationship in the expected direction, with greater strength for PDPs and lesser strength for MAPDs, between the COB measure and OBRAEs.

As described in the methodology, individuals were counted only once in the OBRAE calculation, regardless of how many events they may experience during the measurement year. While this was necessary to align the OBRAE calculation with the measure calculation, it is important to note that small numbers of individuals ('high utilizers') often experience multiple events and drive considerable portions of health care utilization. As a result, this analysis may underestimate the relationship between the COB measure and OBRAEs.

Additionally, this analysis limited the evaluation of OBRAEs to those events resulting in serious health care utilization in the form of inpatient stays and ED visits. Including outpatient claims in the analysis may have increased correlation coefficients by providing more opportunities capture OBRAEs.

Taken as a whole, these findings provide empirical evidence that the COB measure is a valid measure of health plan quality, and a valuable tool that health plans can use to improve the quality of care for their members and decrease the risk that their members will experience negative outcomes.

Although not conducted at the plan-level, PQA strongly recommends panel review of Lo-Ciganic et al (1), described in detail in the evidence form. This research applied PQA's COB measure as specified to a patient-level retrospective cohort methodology, and found that exposure to concurrent use of opioids and benzodiazepines was strongly associated with overdose risk in the subsequent year (adjusted HR 1.82; 95% CI 1.58-2.10), providing compelling further evidence of empirical validity.

1) Lo-Ciganic J, et al. 2018. Geographic Variation of High-Risk Opioid Use and Risk of Overdose Among Disabled Medicare Beneficiaries in the US from 2011 to 2015. Value in Health. 21(2018): S2.

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*)

Patients at end of life, undergoing hospice care, with sickle cell disease, and those with cancer may have unusual requirements for pain management. Thus, these are excluded from these measures whenever data are available.

Testing was performed for the hospice exclusion by identifying the number of members in hospice, where available, and determining the percent of the overall population that would be affected by including patients in hospice care.

Cancer exclusions were identified in the Medicare and Medicaid populations using ICD-9 and ICD-10 codes, depending on the time period of the data (ICD-10 coding began in October 2015). Testing involved identifying the number of exclusions, and determining the percent of the overall population that would be affected by including patients with cancer diagnoses.

Patients with sickle cell disease were identified using ICD-10 codes. Testing involved identifying the number of exclusions, determining the percent of the overall population that would be affected by including patients with sickle cell diagnoses, and determining the impact of the exclusion on measure rates. Testing for the sickle cell disease exclusion was completed using a 2019 sample from a major health plan within the Medicare Advantage, PDP, and commercial lines of business.

The exclusions of hospice, sickle cell disease, and cancer are consistent with the 2016 CDC Guideline for Prescribing Opioids for Chronic Pain, which does not apply to active cancer treatment, palliative care, and end-of life treatment because of the unique therapeutic goals, ethical considerations, opportunities for medical supervision, and balance of risks and benefits with opioid therapy in such care.

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*)

For the Medicare population, after applying the age, continuous enrollment and opioid prescription criteria, the hospice patient exclusions ranged from 0.0% to 27.0% among plan contracts, and the cancer exclusions among plan contracts ranged from 0.0% to 52.8%.

For the Medicaid MAX population, after applying the age, continuous enrollment and opioid prescription criteria, the hospice patient exclusions ranged from 0.0% to 0.5% among plans, and the cancer exclusions among plans ranged from 0.0% to 5.9%.

For the one state-based Medicaid program, only one plan was able to identify 3 hospice patients. The cancer exclusion rate was about 4.5% across the three plans.

For the sickle cell exclusion, within the MAPD population, the denominator before applying the exclusion was 460,439, and the denominator after applying the exclusion was 459,654; the prevalence of the exclusion within the eligible population was <0.01%. Within the PDP population, the denominator before applying the exclusion was 572,868, and the denominator after applying the exclusion was 572,848; the prevalence of the exclusion within the eligible population was <0.01%. Within the commercial population, the denominator before applying the exclusion was 206,685, and the denominator after applying the exclusion was 206,343; the prevalence of the exclusion within the eligible population was <0.01%.

The impact of the exclusion on measure rates was +0.1% (MAPD), +<0.1% (PDP), +0.2% (Commercial).

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*)

The Medicare population shows significant impact of the hospice and cancer exclusions. For the hospice exclusion, up to 27% of beneficiaries in some plan contracts were affected by this exclusion, and the cancer exclusion showed that for some plan contracts, more than half of the population would be affected by this exclusion. Without applying these exclusions, these beneficiaries would be included in the measure. These are significant proportions of the population that could potentially impact the measure rate.

For the Medicaid populations, at the plan level, most of the plans did not identify a lot of hospice patients – therefore, no inferences can be drawn from this exclusion. The cancer exclusion had a higher impact. The results show that in some plans, almost 6% of the population has cancer and would be included in the measure if cancer were not excluded. This is a significant proportion of the population that could potentially impact the measure rate.

Analysis suggests that the sickle cell exclusion does not significantly affect denominator sizes or measure rates. This exclusion is important to the measure's face validity, aligns with clinical guidelines, and was identified as appropriate through PQA's rigorous, consensus-based measure maintenance process.

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

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

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

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

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

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

needed to achieve fair comparisons across measured entities.

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

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

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

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

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

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

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

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

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

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

2b3.9. Results of Risk Stratification Analysis:

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

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

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

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

To assess significant differences in measure rates, the data described in sections 1.5 and 1.6 above were used to calculate the mean, median, standard deviation, and interquartile range for the measure rates for the Medicare and Medicaid (MAX) populations. In addition, the rates were divided into quartiles, and a Student's t-test was used to compare the rates of the plans in the 25th percentile to the rates of the plans in the 75th percentile.

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*)

Tables 6 and 7 show the distribution of the measure rates for the Medicare population. The mean rate was 22.2%, with a median rate of 21.4%, with the lowest plan contract rate at 2.1% and the highest plan contract rate of 44.7%.

Table 6. Variation in Measure Rates – 2015 Medicare Sample

Mean	Median	Standard Deviation
22.2%	21.4%	7.3%

Table 7. Interquartile Range of Measure Rates – 2015 Medicare Sample

Statistic	Value
Minimum	2.1%
25th percentile	17.4%
50th percentile	21.4%
75th percentile	27.3%
Maximum	44.7%
Interquartile Range	9.9%
Student's t-test p-value	<.0001

Tables 8 and 9 show the distribution of the measure rates for the Medicaid MAX population. The mean rate was 3.8%, with a median rate of 2.9%. The lowest plan contract rate was 0.0% and the highest plan contract rate was 18.7%.

Table 8. Variation in Measure Rates – 2008 Medicaid MAX Sample

Mean	Median	Standard Deviation
3.8%	2.9%	3.2%

Table 9. Interquartile Range of Measure Rates – 2008 Medicaid MAX Sample

Statistic	Value
Minimum	0.0%
25th percentile	1.6%
50th percentile	2.9%
75th percentile	5.0%
Maximum	18.7%
Interquartile Range	3.4%
Student's t-test p-value	<.0001

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?*)

For the Medicare population, the measure rates showed significant variation, with a standard deviation of 7.3% and an Interquartile Range of 9.9%. There is a statistically significant difference in measure rates between the top and bottom quartile of the plans included in the testing (*P*< .0001 at alpha = 0.05). This variation shows that there are statistically significant and clinically meaningful differences in rates across plans.

For the Medicaid population, the measure rates showed significant variation, with a standard deviation of 3.2% and an Interquartile Range of 3.4%. There is a statistically significant difference in measure rates between the top and bottom quartile of the plans included in the testing (*P*< .0001 at alpha = 0.05). This variation shows that there are statistically significant and clinically meaningful differences in rates across plans.

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

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

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

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

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

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

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

With the use of prescription claims data as the data source for this measure, the dispensing information (including medication, days' supply, quantity dispensed, and dosage) is available for each patient.

Since each of these data elements are available via prescription claims data, it is not expected—nor was it found—that missing data would result. Age is derived from the date of birth in the enrollment data. The date of birth in the CMS Medicare Beneficiaries Summary Files (MBSF) and Medicaid administrative data is considered to largely be valid and reliable since it determines eligibility for enrollment and payment of services.

Patients in hospice are excluded from this measure. No testing was performed on this exclusion as the data source, prescription claims data, do not contain claims for palliative medication, such as opioids, for persons in Medicare Part D that are in hospice care. For the Medicaid population, the majority of the plans were not able to identify hospice exclusions in their data.

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*)

No missing data was found in the testing of this measure.

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

As stated above, no missing data was found through testing, nor would missing data be expected to occur in the future. Therefore, performance results would not be biased, as prescription claims data provides the data elements necessary to calculate the measure rate.

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), Other

If other: Medical claims data, Prescription claims data, Enrollment Data

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).

---UPDATED FOR MAINTENANCE---

PQA does not intend to develop an electronic clinical quality measure (eCQM) version of this health plan claims-based performance measure. However, PQA is currently exploring opportunities to convert existing claims-based measures to a digital clinical quality measure (dCQM) format, to align with CMS' stated goal of using dCQMs by 2025.

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.

Pilot test sites indicated the measure was feasible and results were able to be reported efficiently, accurately, and without difficulty. The required data (prescription claims and medical claims) are readily available.

---UPDATED FOR MAINTENANCE---

PQA is not aware of difficulties in implementing the measure into the programs described in 4.1. The measure is specified using prescription and medical claims data, which are readily available and accurate.

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)*.

PQA retains the rights to the measures and can rescind or alter the measures at any time. PQA may approve an organization's use of the measures; however, no organization may use the measures without first obtaining permission from PQA prior to using the measures. Certain uses of the measures are only approved with a licensing agreement from PQA that specifies the terms of use and the licensing fee. PQA reserves the right to determine the conditions under which it will approve use and/or license the measures. Users of the measures shall not have the right to alter, enhance, or otherwise modify the measures.

National Drug Code and ICD code value sets are required to calculate the measure and are provided with the narrative specifications to licensees.

---UPDATED FOR MAINTENANCE---

All uses of PQA Measures are subject to such conditions as PQA specifies, and will be subject to a license agreement specifying the terms of use and the license fee. Government agencies do not pay a license royalty.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
*	Public Reporting Medicare Part D Display Page https://www.cms.gov/files/zip/2021-display-measures.zip Quality Improvement (external benchmarking to organizations) Medicare Part D Patient Safety Reports https://www.hhs.gov/guidance/sites/default/files/hhs-guidance-documents/april%25202019%2520patientsafety_oms_updates_2.pdf Medicaid Adult Core Set https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-core-set/index.html

*cell intentionally left blank

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

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

Program name & sponsor: Centers for Medicare & Medicaid Services (CMS) Medicaid Adult Core Measure Set. National program with state-level voluntary reporting.
Purpose: The Affordable Care Act (Section 1139B) requires the Secretary of Health & Human Services (HHS) to identify and publish a core set of health care quality measures for adult Medicaid enrollees. The core set is published for voluntary use by state Medicaid programs. State data derived from the core measures are part of CMS's annual Child and Adult Core Set measure reporting, which includes publication of datasets that highlight publicly reportable measures. CMS annually releases information on state progress in reporting the Adult Core Set measures and assesses state-specific performance for measures that are reported by at least 25 states and which met internal standards of data quality.
Geographic area: This is a national program with state-level reporting.
Level of measurement and setting: Health plan level of measurement. Outpatient setting.
---UPDATED FOR MAINTENANCE---
Name of program and sponsor: Medicare Part D Display page
Purpose: The Part D Display page is a public reporting program, which include measures that have been transitioned from the Star Ratings, new measures that are tested before inclusion into the Star Ratings, or measures displayed for informational purposes only. Display measures are available to the public via the Part C and D Performance Data [https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData] page.
Geographic area and number and percentage of accountable entities and patients included: Measure rates are reported for all contracts in the Medicare Part D program meeting the minimum denominator requirement of 30, except for contracts whose measurement period is prior to one year past the contract's effective date, in which case the contract is marked as "Plan too new to be measured", or contracts that were otherwise not required to report the measure to CMS.
Level of measurement and setting: Health plan level of measurement. Outpatient setting.

Name of program and sponsor: Medicare Part D Patient Safety Reports
Purpose: The Medicare Part D Patient Safety Reports are made available to all Part D plan sponsors on a monthly basis for the purposes of quality improvement, allowing plan sponsors to monitor their performance on key quality metrics and compare their performance to program averages.
Geographic area and number and percentage of accountable entities and patients included: This is a national program representing performance by plan sponsors spanning the full Medicare Part D program. In 2018, per the Chronic Conditions warehouse, approximately 46.7 million patients were enrolled in the Medicare Part D program across more than 650 sponsors (refer to 1b.2)
Level of measurement and setting: Health plan level of measurement. Outpatient setting.

Program name & sponsor: Centers for Medicare & Medicaid Services (CMS) Medicaid Adult Core Set. National program with state-level voluntary reporting.
Purpose: The Affordable Care Act (Section 1139B) requires the Secretary of Health & Human Services (HHS) to identify and publish a core set of health care quality measures for adult Medicaid enrollees. The core set is published for voluntary use by state Medicaid programs. State data derived from the core measures are part of CMS's annual Child and Adult Core Set measure reporting, which includes publication of datasets that highlight publicly reportable measures.
Geographic area: This is a national program with state-level reporting.
Level of measurement and setting: Health plan level of measurement. Outpatient setting.
** Please note that with regard to the Medicaid Adult Core Set, the Concurrent Use of Opioids and Benzodiazepines was used in 22 states as of the most recent available data (FFY2019). Beginning in FY2024, states will be required to report on the core set of behavioral health measures for adults enrolled in Medicaid, including the Concurrent Use of Opioids and Benzodiazepines measure.

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?*)
This is a new measure that was developed in 2016. The measure was added to the 2018 Medicaid Adult Core Measure Set; however, measures in the program are publicly reported only if 25 or more states report on the measure. Given that 2018 is the first year the measure is included, it is not yet publicly reported. We would anticipate adoption of the measure over time, with public reporting once 25 or more states are reporting on the measure.

---UPDATED FOR MAINTENANCE---

N/A. At time of maintenance, the measure has been implemented in the Medicaid Adult Core Set, the Medicare Part D Display page, and the Medicare Part D Patient Safety Reports.

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.*)

The measure has been added to the Medicaid Adult Core Set for 2018. CMS annually releases information on state progress in reporting the Adult Core Set measures and assesses state-specific performance for measures that are reported by at least 25 states and which met internal standards of data quality.

PQA not only develops and stewards its measures, it also dedicates resources to outreach and implementation efforts. PQA disseminates information regarding the availability of its measures, and provides technical assistance to those implementing or considering implementing PQA-endorsed measures.

Additionally, per the CMS Advance Notice of Methodological Changes for Calendar Year 2019 for Medicare Advantage Capitation Rates, Part C and Part D Payment Policies and 2019 Draft Call Letter (available: https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/Advance2019Part2.pdf), CMS proposes to begin reporting the Concurrent Use of Opioids and Benzodiazepines measure in the Medicare Part D Patient Safety reports for the 2018 measurement year, and to add it to the Medicare Part D display page for 2021 (using 2019 data) and 2022 (using 2020 data). CMS also will consider this measure for the 2023 Star Ratings (using 2021 data) pending rulemaking.

---UPDATED FOR MAINTENANCE---

N/A. At time of maintenance, the measure has been implemented in the Medicaid Adult Core Set, the Medicare Part D Display page, and the Medicare Part D Patient Safety Reports.

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.

This measure has been reported to Medicare Part D plan sponsors since 2018 in the monthly Patient Safety reports via CMS' vendor, Acumen. The measure has also been in the Medicaid Adult Core set since 2018 with technical assistance provided by CMS' vendor Mathematica. As measure steward, PQA provides timely technical assistance to questions sent via email to an email box, as well as questions received via phone or other channels.

---UPDATED FOR MAINTENANCE---

PQA uses a transparent, consensus-driven process to draft, test, refine, endorse, and maintain quality measures. For more information on the multiple stakeholder groups involved in the original development of this measure, their feedback, and their assessments of face validity, please refer to section 2b.1.1 of the testing form.

PQA not only develops and stewards its measures, it also dedicates resources to outreach and implementation efforts. PQA disseminates information regarding the availability of its measures and provides technical assistance to those implementing or considering implementing PQA measures, including both program administrators and, in some cases, individual entities participating in programs.

Within the Medicare Part D Patient Safety Reports, Part D plan sponsors (including MAPDs and PDPs) may download and review their measure performance each month via a measure package. These actionable measure packages include a summary contract-level report for each measure and additional beneficiary-level files. Part D plan sponsors can use the Patient Safety Reports to compare their performance to overall averages and monitor their progress in improving their measure rates. CMS provides memorandums educating Part D plan sponsors on the Patient Safety Reports program, what measures are included, how to access their results, and more. For more information, refer to the UPDATES – 2020 Medicare Part D Patient Safety Reports [https://www.hhs.gov/guidance/sites/default/files/hhs-guidance-documents/april%202020%20patient%20safety%20updates_8.pdf] memorandum.

Within the Medicare Part D Display page, data are publicly available via the Part C and D Performance [https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData] page. Part D plan sponsors can use Display data to compare their performance to other plans and national averages and monitor their progress in improving their measure rates. CMS provides memorandums educating Part D plan sponsors on the Display measures, what measures are included, how to access their results, and more. For more information, refer to the Plan Preview of Display Measures in HPMS [https://www.hhs.gov/guidance/sites/default/files/hhs-guidance-documents/display%20measures%20memo_2018_2_0.pdf] memorandum and the 2021 Display Measure Technical Notes, available from the Part C and D Performance Data [https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData] page.

For the Medicaid Adult Core Set, states calculate measure rates for submission to the Core Set. Twenty two state Medicaid programs are currently reporting the measure and are aware of their own performance. All 50 states are required to report on the Concurrent Use of Opioids and Benzodiazepines measure by FFY2024.

Additionally, to increase the number of states consistently collecting, reporting, and using the Medicaid Adult Core Set measures, CMS established the Technical Assistance and Analytic Support (TA/AS) Program with an award of a contract to Mathematica Policy Research. Mathematica – teamed with the National Committee for Quality Assurance and the Center for Health Care Strategies – works with CMS to support states' adult health care quality measurement and improvement efforts. The overarching goals are to increase the number of states consistently collecting and uniformly reporting the voluntary core measures set, and to help states understand how to use these data to improve the quality of care for adults. As part of the technical assistance effort, CMS will share promising practices for collecting the core measures with states. For more information, refer to Medicaid.gov [https://www.medicaid.gov/medicaid/quality-of-care/performance-measurement/adult-and-child-health-care-quality-measures/adult-core-set-reporting-resources/index.html].

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.

This measure has been reported to Medicare Part D plan sponsors since 2018 in the monthly Patient Safety reports via CMS' vendor, Acumen. The measure has also been in the Medicaid Adult Core set since 2018 with technical assistance provided by CMS' vendor Mathematica. As measure steward, PQA provides timely technical assistance to questions sent via email to an email box, as well as questions received via phone or other channels.

---UPDATED FOR MAINTENANCE---

Section 4a2.1.1 provides detail on reporting provided by each program. PQA does not collect data or report measure scores; however, as the measure steward, PQA provides technical assistance to support accurate

implementation of the measure specifications and may also provide educational webinars or other assistance as necessary to support measure implementation.

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.

---UPDATED FOR MAINTENANCE---

As a measure steward, PQA disseminates information regarding the availability of its measures, and provides technical assistance to those implementing or considering implementing PQA measures. PQA also considers feedback provided via technical assistance and other avenues for potential measure updates. For more information on the multiple stakeholder groups involved in the original development of this measure, their feedback, and their assessments of face validity, please refer to section 2b.1.1 of the testing form.

PQA receives feedback from measure users via a web form [https://www.pqaalliance.org/tech-assist-form]. PQA staff then provide timely (i.e., 24-48 hours) responses to all inquiries by email, telephone or webinar. Technical assistance questions are regularly reviewed for opportunities to clarity and refine PQA measures through PQA's consensus-based measure update process.

Additionally, CMS shares all comments related to PQA measures included in Part D quality programs -- including those specific to the Concurrent Use of Opioids and Benzodiazepines measure -- that they receive in response to proposed rules and the Part D draft Call Letter, which are released on an annual basis. Comments are reviewed by PQA staff and brought to the MUP which then determines whether refinements or clarifications to the specifications are needed.

Additionally, within the Medicaid Adult Core Set, PQA sends updated versions of measure specifications to the program contractor on a regular basis and takes into consideration any comments or suggestions received through the program contractor.

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

---UPDATED FOR MAINTENANCE---

Since initial development and endorsement, PQA received feedback that an exclusion for patients with a sickle cell disease diagnosis is appropriate for the measure from a variety of sources. After soliciting expert input and completing review, these recommendations were presented to and approved by PQA's Measure Update Panel, and PQA's Quality Metrics Expert Panel. Individuals with sickle cell disease have unique pain management needs, and the CDC have stated that their Guideline for Prescribing Opioids for Chronic Pain is not intended to apply to patients with sickle cell disease [Available at https://www.asco.org/sites/new-www.asco.org/files/content-files/advocacy-and-policy/documents/2019-CDC-Opioid-Guideline-Clarification-Letter-to-ASCO-ASH-NCCN.pdf]. Due to these considerations, and their unique therapeutic goals, ethical considerations, opportunities for medical supervision, and balance of risks and benefits, individuals with a diagnosis of sickle cell disease are excluded from this measure.

PQA has received feedback from measure users suggesting that a palliative care exclusion may be appropriate for the measure. Patients receiving palliative care have unique therapeutic goals, and the balance of risks and benefits associated with opioid use may be different from the broader population. As a result, PQA is evaluating the appropriateness of adding a palliative care exclusion to the measure through our standard, consensus-based measure update process.

Additionally, PQA has received feedback from measure users suggesting that an exclusion for patients in long-term care may be appropriate for the measure. The CMS Part D Opioid Safety Edits in the Overutilization Monitoring System, including the soft-edit for concurrent opioid and benzodiazepine use, exclude persons who are residents of a long-term care facility, and individuals in long-term care are monitored more closely than the general population, resulting in different opioid-related risks. As a result, PQA is evaluating the appropriateness of adding a long-term care exclusion to the measure through our standard, consensus-based maintenance process.

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

N/A

---UPDATED FOR MAINTENANCE---

Not applicable.

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

N/A

---UPDATED FOR MAINTENANCE---

As noted in section 4a2.2.2, the exclusion of individuals with a sickle cell disease diagnosis was added to the measure. Potential exclusions for patients receiving palliative care and long-term care are currently being evaluated through PQA's standard, consensus-based measure update process.

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.

The performance results can be used to establish benchmarks and identify opportunities to decrease co-prescribing of opioid and benzodiazepines. Sun et al. estimated that the elimination of the concurrent use of opioids and benzodiazepines could reduce the population risk of an emergency department visit or hospital admission for opioid overdose by 15%.(1) Despite the risks, concurrent prescriptions for opioids and benzodiazepines are relatively common and increasing. From 2001-2013, concurrent prescribing increased by nearly 80% (from 9% to 17%) among privately insured patients.(1) In one study, approximately half of the patients received both opioid and benzodiazepine prescriptions from the same prescriber on the same day.(2) In a 2015 analysis of Medicare Part D non-cancer and/or non-hospice patients on opioid therapy, the prevalence of benzodiazepine concurrent use was 24%.(3)

1. Sun EC, Dixit A, Humphreys K, et al. Association between concurrent use of prescription opioids and benzodiazepines and overdose: retrospective analysis. BMJ. 2017;356:j760. doi: 10.1136/bmj.j760. PMID: 28292769

2. Hwang CS, Kang EM, Kornegay CJ, Staffa JA, Jones CM, McAninch JK. Trends in the Concomitant Prescribing of Opioids and Benzodiazepines, 2002-2014. Am J Prev Med. 2016:1-10. doi:10.1016/j.amepre.2016.02.014.

3. CMS. Concurrent Use of Opioids and Benzodiazepines in a Medicare Part D Population. May 12, 2016. 2016. https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Concurrent-Use-of-Opioids-and-Benzodiazepines-in-a-Medicare-Part-D-Population-CY-2015.pdf. Accessed December 6, 2016.

---UPDATED FOR MAINTENANCE---

Data from 2018 and 2019 in the Medicare Part D Patient Safety Reports demonstrate a downward trend across both the MAPD and PDP lines of business:

MAPD Mean 2018: 19.44%

MAPD Mean 2019: 17.39%

Trend: -2.05%

PDP Mean 2018: 19.36%

PDP Mean 2019: 17.44%

Trend: -1.92%

In addition to demonstrating a promising trend of improving performance, the performance distributions provided in 1b.2. demonstrate significant variation and room for improvement. Standard deviations range from 3.98% to 6.72% depending on the line of business and year. Additionally, upper deciles and maximum rates suggest that certain contracts have substantially higher rates than the mean. In these cases, the Medicare Part D Patient Safety Reports are a critical tool for plan sponsors to monitor their performance and improve over time. PQA will continue to monitor data from the Medicare Part D Patient Safety Reports as they become available, to evaluate the continued trends in plan sponsor performance.

--

Data from the Medicare Part D Display page were only available for calendar year 2019, which does not allow for trend analysis. However, given that the measured population is nearly identical to the Medicare Part D Patient Safety reports as noted in section 1b.2, those data can be referenced to gain insight in COB measure rate trends in the Medicare Part D Display page.

Data for the Medicaid Adult Core Set were only available for calendar year 2018, which does not allow for trend analysis. As more states report on the measure, PQA will continue to monitor trends for improvement. However, the current distribution demonstrates significant variation and room for improvement. The standard deviation was 5.36%, and upper deciles and maximums suggest that certain states have substantially higher rates than the mean (19.15%). Given this variation, the COB measure serves as a valuable quality improvement tool for states to monitor their performance and improve their rates over time.

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

---UPDATED FOR MAINTENANCE---

As a steward, PQA provides technical assistance for our measures in various implementations. In technical assistance requests and in discussion with program stewards, PQA has not identified unintended consequences or unexpected findings associated with this measure.

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

---UPDATED FOR MAINTENANCE---

Not applicable.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

2940 : Use of Opioids at High Dosage in Persons Without Cancer

2950 : Use of Opioids from Multiple Providers in Persons Without Cancer

2951 : Use of Opioids from Multiple Providers and at High Dosage in Persons Without Cancer

3316 : Safe Use of Opioids – Concurrent Prescribing

3541 : Annual Monitoring for Persons on Long-Term Opioid Therapy (AMO)

3558 : Initial Opioid Prescribing for Long Duration (IOP-LD)

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

Related measures:

• Use of Opioids at High Dosage (NCQA)

• Use of Opioids from Multiple Providers (NCQA)

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

Yes

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

---UPDATED FOR MAINTENANCE--- At time of maintenance, PQA has also identified the 3316e: Safe Use of Opioids – Concurrent Prescribing measure as related. Although the area of focus overlaps, 3316e is specified at the facility level as an eCQM, as opposed to 3389, which is specified at the health plan level and is claims-based. PQA identified the 3558: Initial Opioid Prescribing for Long Duration and 3541: Annual Monitoring for Persons on Long-Term Opioid Therapy) measures as related to opioid prescribing, although the areas of focus (initial opioid prescribing and annual monitoring) are different than 3389 (concurrent use of opioids and benzodiazepines).

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.)

There are no competing measures (i.e., those that addresses both the same measure focus and the same target population).

---UPDATED FOR MAINTENANCE---

There are no competing measures (i.e., those that addresses both the same measure focus and the same target population).

Appendix

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

No appendix **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): PQA, Inc.

Co.2 Point of Contact: Ben, Shirley, bshirley@pqaalliance.org, 703-347-7938-

Co.3 Measure Developer if different from Measure Steward: PQA, Inc.

Co.4 Point of Contact: Ben, Shirley, bshirley@pqaalliance.org, 703-347-7938-

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.

347A diverse group of stakeholders, including health plans and PBMs (those organizations that will be measured) were well represented throughout the entire development process, including contributing to defining the specifications as members of the Measure Development Team, as testers using the measure specifications to calculate the rates, in the review for face validity and review of testing results as members of the Quality Metrics Expert Panel, and in the vote for PQA endorsement.

PQA Measure Development Teams are small, technically proficient teams composed of diverse stakeholders, to develop individual metrics. Measure Development Team 13 (MDT 13) developed the Concurrent Use of Opioids and Benzodiazepines measure. The members of MDT 13 and their corresponding organizations are listed below:

Cyndi Barham, PharmMD

Maribeth Bettarelli, CVS Health

Donna Boreen, BCBSMN

Jeffrey Bratberg, University of Rhode Island College of Pharmacy (representing the American Association of Colleges of Pharmacy)

Sara Burnheimer, UPMC Health Plan

Pauline Chan, California Department of Health Care Services

Alexandra Cruz, Healthfirst

Samuel Currie, Horizon NJ Health (representing the Association for Community Affiliated Plans)

Tiffany Del Rosario, SCAN Health Plan

Angela DeVeaugh-Geiss, Purdue Pharma LP

Jeff Fink, Express-Scripts

Rainelle Gaddy, Humana

Travis Gau, Medication Management Solutions

Adriane Irwin, American Association of Colleges of Pharmacy

Shellie Keast, University of Oklahoma

Richard Logan, MedHere Today

Michael Long, APhA

Denis Matsuoka, Kaiser Permanente

Karen McLin, SinfoniaRx

Alina Meile, Aetna

Mary Miller, Rite Aid

Anna Polk, Centers for Medicare & Medicaid Services

Madeline Ritchie, Aetna (representing the Academy of Managed Care Pharmacy)

Jennifer Shin, OptumRx

Mindy Smith, PrescribeWellness

Jennifer Snyders, Cigna-HealthSpring

Kathleen Vest, Midwestern University Chicago College of Pharmacy

PQA´s Measure Validity Panel (MVP) is a small group of individuals appointed by PQA staff, to determine whether the performance scores resulting from the measure can be used to distinguish good from poor quality clinical care (i.e., validity). The MVP members that reviewed the Concurrent Use of Opioids and Benzodiazepines measure and their corresponding organizations are listed below:

Susan Skledar, University of Pittsburgh

Ben Banahan, University of Mississippi

Jeff Pohler, University of FL College of Pharmacy

Dan Rehrauer, HealthPartners

Kyle Null, Takeda

Marybeth Farquhar, URAC

PQA´s Quality Metrics Expert Panel (QMEP) is a small group of individuals, selected by PQA staff through an application process, to recommend measure concepts for testing, review measure testing results, and recommend measures for endorsement consideration by PQA membership. The QMEP members that reviewed the Concurrent Use of Opioids and Benzodiazepines measure and their corresponding organizations are listed below:

Amanda Brummel, Fairview Health Services

Bimal Patel, MedImpact

Catherine Coast, Highmark

Christopher Dezii, Bristol-Myers Squibb Company

Christopher Powers, Centers for Medicare & Medicaid Services

Craig Schilling, Optum

David Nau, Pharmacy Quality Solutions

Gary Erwin, Omnicare

Jenny Weber, Humana

Jessica Frank, OutcomesMTM

Karen Farris, University of Michigan College of Pharmacy

Keith Widmer, Express Scripts

Kent Summers, Astellas

Lynn Deguzman, Kaiser Permanente, Northern California

Mary Ann Kliethermes, Midwestern University

Mitzi Wasik, Aetna

Pat Gleason, Prime Therapeutics

Steven Riddle, Wolters Kluwer Health

Steven Burch, GlaxoSmithKline

Tony Willoughby, McKesson

Tripp Logan, Logan and Seiler

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2016

Ad.3 Month and Year of most recent revision: 02, 2021

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

Ad.5 When is the next scheduled review/update for this measure? 02, 2022

Ad.6 Copyright statement: COPYRIGHT 2021 PQA, INC. ALL RIGHTS RESERVED

Ad.7 Disclaimers: This measure is not intended for clinical-decision-making. This measure is intended for retrospective evaluation of populations of patients and should not be used to guide clinical decisions for individual patients. For clinical guidance on opioid prescribing, see the Center for Disease Control and Prevention CDC Guideline for Prescribing Opioids for Chronic Pain [https://www.cdc.gov/mmwr/volumes/65/rr/rr6501e1.htm?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fmmwr%2Fvolumes%2F65%2Frr%2Frr6501e1er.htm] and Guideline Resources [https://www.cdc.gov/drugoverdose/prescribing/resources.html].

Ad.8 Additional Information/Comments: N/A