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 #: 3558

Corresponding Measures:

De.2. Measure Title: Initial Opioid Prescribing for Long Duration (IOP-LD)

Co.1.1. Measure Steward: Pharmacy Quality Alliance

De.3. Brief Description of Measure: The percentage of individuals 18 years of age and older with one or more initial opioid prescriptions for >7 cumulative days' supply.

1b.1. Developer Rationale: Opioid misuse, addiction, and overdose are a public health crisis affecting social and economic welfare in the United States, with more than 130 Americans dying each day due to opioid overdose.(1) Although recent data suggest slight decreases in overdose deaths involving natural/semisynthetic opioids and methadone, overdose deaths involving synthetic opioids continue to grow, and the rates across all types of opioids remain unacceptably high.(2)

Prescription opioids for pain management remain a contributing factor to the crisis. A systematic review of 38 studies, conducted in 2015, found that approximately 21% to 29% of patients prescribed opioids for chronic pain misuse them, while the 2015 National Survey on Drug Use and Health found that 12.5% of adults with opioid prescriptions reported misuse, and 16.7% reported a prescription opioid use disorder.(3,4) In response to the opioid overdose epidemic, a public health emergency was declared in 2017 by the United States Department of Health and Human Services.(5)

The duration of initial opioid exposure is associated with a higher likelihood for high-risk and long-term opioid use, misuse, overdose, and other negative outcomes (6,7,8,9,10). The 2016 Centers for Disease Control and Prevention (CDC) Guideline for Prescribing Opioids for Chronic Pain recommends that when opioids are used for acute pain (i.e., pain with abrupt onset and caused by an injury or other process that is not ongoing), no greater quantity should be prescribed than is needed for the expected duration of pain severe enough to require opioids.(11) According to the guideline, three days or less will often be sufficient, and more than seven days of opioids will rarely be needed for the treatment of acute pain.

Published studies support the CDC recommendations to limit the duration of initial opioid use. In 2017, Shah and colleagues published a retrospective cohort study using claims data from a nationally representative database of commercially insured patients, examining the relationship between initial opioid prescription characteristics and the likelihood of opioid discontinuation.(6) Increasing days' supply of the first prescription was consistently associated with a lower likelihood of opioid discontinuation: 3-4 days' supply (Hazard Ratio

[HR], 0.70; 95% Confidence Interval [CI], 0.70-0.71); 5-7 days' supply (HR, 0.48; 95% CI, 0.47-0.48); 8-10 days' supply (HR, 0.37; 95% CI, 0.37-0.38); 11-14 days' supply (HR, 0.32; 95% CI, 0.31-0.33).

Zhang and colleagues found that an initial opioid prescription of greater than seven days' supply was associated with a significant ($p<0.001$) increase in high-risk opioid use, including overlapping opioid prescriptions, concurrent use of opioids and benzodiazepines, and the probability of three or more opioid prescribers, as well as increased likelihood of long-term opioid use.(7) In addition, Brat and colleagues published a study in 2018 evaluating the effects of varying opioid prescribing patterns after surgery on misuse or overdose in a retrospective cohort study of an opioid naïve population. (8) The total duration of opioid use was the strongest predictor of misuse, with an estimated increase rate of 20% of opioid misuse per each week of an opioid prescription after adjusting for covariates. A study by Durand et al of 46,399 opioid-naive injured workers found in comparison to a reference group of <5 days' supply, an initial prescription with 5-9 days' supply was associated with a significant increase in the odds of long-term use (adjusted OR 1.83, 95% CI 1.56-2.14).(9) Hadlandsmyth and colleagues found that days' supply was the strongest predictor in a multivariable model of long-term opioid use, with initial opioid prescriptions for 7 days or fewer serving as a reference group, and greater days' supply associated with increased risk of long-term opioid use, including 8-14 days (adjusted OR 1.44, 95% CI 1.38-1.51).(10) For a more detailed review of the evidence, please refer to the evidence form.

Aligned with CDC recommendations and published evidence, this performance measure evaluates the percentage of individuals 18 years of age and older with one or more initial opioid prescriptions for >7 cumulative days' supply. Patients with cancer diagnoses, sickle cell diagnoses, and those receiving hospice care are excluded from the measure because of their unique therapeutic goals, ethical considerations, opportunities for medical supervision, and balance of risks and benefits.(11, 12) This measure was designed for retrospectively evaluating health plan performance at the population level and is not intended to guide clinical care for individual patients.

Citations:

1) NIH/NIDA. Opioid Overdose Crisis [Internet]. 2019 [cited 2019 Mar 29]. Available from: https://www.drugabuse.gov/drugs-abuse/opioids/opioid-overdose-crisis.

2) Hedegaard H, Minino AM. Warner M Drug Overdose Deaths in the United States, 1999-2018. NCHS Data Brief No 365. January 2020. Available from: https://www.cdc.gov/nchs/products/databriefs/db356.htm

3) Han B, Compton WM, Blanco C, et al. Prescription Opioid Use, Misuse, and Use Disorders in U.S. Adults: 2015 National Survey on Drug Use and Health. Ann Intern Med. 2017;167(5):293-301. PMID: 28761945.

4) Vowles KE, McEntee ML, Julnes PS, Frohe T, Ney JP, van der Goes DN. Rates of opioid misuse, abuse, and addiction in chronic pain: a systematic review and data synthesis. Pain. 2015; 156:569-76. PMID: 25785523.

5) HHS. National Rx Drug Abuse and Heroin Summit. Secretary Price Announces HHS Strategy for Fight Opioid Crisis. 2017. Available from: https://www.hhs.gov/about/leadership/secretary/speeches/2017-speeches/secretary-price-announces-hhs-strategy-for-fighting-opioid-crisis/index.html.

6) Shah A, Hayes CJ, Martin BC. Factors Influencing Long-Term Opioid Use Among Opioid Naive Patients: An Examination of Initial Prescription Characteristics and Pain Etiologies. J Pain. 2017; 18:1374-83. PMID: 28711636.

7) Zhang Y, Johnson P, Jeng PJ, et al. First Opioid Prescription and Subsequent High-Risk Opioid Use: a National Study of Privately Insured and Medicare Advantage Adults. J Gen Intern Med. 2018;33(12):2156-2162. PMID: 30206790.

8) Brat GA, Agniel D, Beam A, et al. Postsurgical prescriptions for opioid naive patients and association with overdose and misuse: retrospective cohort study. BMJ. 2018; 360:j5790. PMID: 29343479.

9) Durand, Z., Nechuta, S., Krishnaswami, S., Hurwitz, E.L. and McPheeters, M., 2019. Prevalence and Risk Factors Associated With Long-term Opioid Use After Injury Among Previously Opioid-Free Workers. JAMA network open, 2(7), pp.e197222-e197222. PMID: 31314119.

10) Hadlandsmyth K, Lund BC, Mosher HJ. Associations between initial opioid exposure and the likelihood for long-term use. J Am Pharm Assoc . 2019;59(1):17-22. PMID: 30409501.

11) Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016;65:1-49. PMID: 26987082.

12) Carlon RW, Hudis CA, Ligget M. 2019 CDC Opioid Guideline Clarification Letter to ASCO / ASH / NCCN. 2019 Feb. Centers for Disease Control and Prevention. Available from 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.

S.4. Numerator Statement: The number of individuals from the denominator with >7 cumulative days' supply for all opioid prescription claims within any opioid initiation period.

The opioid initiation period is defined as the date of service of the initial opioid prescription plus two days, i.e., the 3-day time period when the numerator is assessed.

S.6. Denominator Statement: The denominator includes individuals 18 years of age or older with one or more prescription claims for an opioid and a negative medication history for any opioid medication during the 90-day lookback period.

S.8. Denominator Exclusions: Individuals with cancer, sickle cell disease, or in hospice at any point during the measurement year or the 90 days prior to the first day of 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: Most Recent Endorsement Date:

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? Not applicable.

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

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

The developer provides the following evidence for this measure:

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

Evidence Summary

- The developer cited the Centers for Disease Control and Prevention (CDC) Guideline for Prescribing Opioids for Chronic Pain, which recommends that initial prescription of opioids for the treatment of acute pain should be of the lowest effective dose of immediate-release and must not exceed seven days' supply (Recommendation Category: A, evidence type: 4).
- Recommendation = Category A: apply to all persons in a specified group and indicate that most patients should receive the recommended course of action. Evidence = Type 4: clinical experience and observations, observational studies with important limitations, or randomized clinical trials with several major limitations.
- The developer cited systematic reviews of retrospective cohort studies consistent with the CDC's guidelines thus reinforcing the relationship between initial opioid prescription and the risk of long-term use and a variety of adverse health outcomes.

Questions for the Committee:

- *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?*
- *If derived from patient report, does the target population value the measured process or structure and find it meaningful?*

Guidance from the Evidence Algorithm

Process measure based on systematic review (Box 3) → QQC Provided (Box 4) → Quantity: Moderate; Quality: Moderate; Consistency: Moderate (Box 5b) → Moderate

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer used administrative claims data for Medicare, Medicaid, and commercial populations.
    - Medicaid data from 2017 and included nine health plans with a total of 728,645 individuals aged 18 and older.
        - Plan/contract-level rates ranged from 9.5% to 33.5%
        - Mean rate: 23.7%, Median rate: 25.9%, and Standard deviation: 8.1%
    - Medicare data from 2018 and included 673 Medicare Advantage Prescription Drug (MAPD) plans and 61 standalone Prescription Drug Plans (PDPs) with a total of 46,675,451 individuals aged 18 and older.
        - Plan/contract-level rates ranged from 16.7% to 86.6%
        - Mean rate: 43.8%, Median rate: 41.8%, and Standard deviation: 11.9%
    - Commercial data from 2017 and included a total of three health plans across three states with a total of 1,266,256 individuals aged 18 and older.
        - Plan/contract-level rates ranged from 23.7% to 26.8%, Median Rate: 24.7%, and Standard deviation: 1.6%.

- The developer indicated that Medicare and Medicaid data sets had statistically significant differences between performance in the 25th and 75th percentiles. The developer did not perform a Student's T-test on commercial plans due to insufficient plans in the commercial sample.

Disparities

- The developer indicated that for the Medicare data, measure rates were 36.4% among males and 43.6% among females. For age bands, measure rates were 30.2% in 18-50, 45.1% in 51-64, 39.5% in 65-85, and 52.7% in 85+. Measure rates were 44.7% among beneficiaries with low-income subsidy status (LIS), and 39.1% among non-LIS.
- For the Medicaid data, measure rates were 19.9% among males and 14.7% among females. For age bands, measure rates were 11.5% in 18-50, 30.8% in 51-64, 52.0% in 65-84, and 58.1% in 85+.
- For the commercial data, measure rates were 26.9% among males and 25.8% among females. For age bands, measure rates were 21.5% in 18-50, 34.2% in 51-64, 36.5% in 65-84, and 58.8% among 85+.
- The developer provided additional literature citing higher odds of long-term opioid use in rural residences compared with urban residences (OR, 1.51; 95% CI 1.31-1.73) and literature citing disparities in opioid-related outcomes, such as overdose.

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

Comments:

**This metric addresses a well defined risk factor for opioid dependence. Applies directly and is conceptionally sound.

**I would call the relationship of process to outcome as low, not moderate since the relationship between shorter initial prescriotion and opioid overuse is correlation not causation.

**No

**Developers provide convincing evidence that long duration opioid prescriptions is a contributing factor to the opioid crisis.

**The evidence to support this measure is strong.

**Moderate – Process measure logic: Patient visits provider presenting with acute pain Provider prescribes opioids to manage acute pain with greater than 7 days' supply Patient is at increased risk of long-term opioid use, opioid misuse, and overdose. Based on systematic review consistent with CDC Guideline for Prescribing Opioids for Chronic Pain. Recommendation = Category A: apply to all persons in a specified group and indicate that most patients should receive the recommended course of action. Evidence = Type 4: clinical experience and observations, observational studies with important limitations, or randomized clinical trials with several major limitations.

**Not sure the measure directly relates to what it purporting to show

1b. Performance Gap

Comments:

**yes, increases in prescription supply

**It is not totally clear if the rates provided are the percentage of patients with >7d initial prescription, if that is true, there seems to be enough variability to support a quality metric. Descriptive data on the distribution especially by a random sample of providers would be helpful.

**Large opportunity amongst many settings and groups

**The developers provide medicaid, medicare and commercial data to successfully demonstrate a performance gap.

**Performance gap was clearly demonstrated with Medicare and Medicaid data. However, I am concerned that the gap was less evident with commercial one, which contributes to a large portion of the health data.

**High – Rates vary by lines of business: Medicaid (9.5% to 33.5%), Medicare (16.7% to 86.6%), and Commercial (23.7% to 26.8%). Medicare and Medicaid data statistically significant differences at 25th to 75th percentiles.

**There was a gap shown but co-factors were not presented

Disparities:

**Age, sex, socioeconomic, rural

**Disparities by gender have been reviewed, and with low-income subsidy and rural areas. There is some disparity by the LIS

**See above - well noted

**The developers provide medicaid and medicare data to demonstrate disparities in the care related to long duration opioid prescriptions

**In the developer's analysis, disparities were shown when comparing among different age groups, with elderly age 85+ having a consistent higher measure rate among Medicare and Medicaid patients. Disparity was also evident for Medicare beneficiaries with low-income subsidy status.

**No

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The staff or is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The staff is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

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

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

2a1. Reliability – Specifications
Comments:

**Should be reliable, easily obtained data

**The reliability to measure days of prescription >7 days seems quite reliable. The measurement of "overprescription is less reliable. Some high risk conditions are excluded and only first prescriptions are measured which increases the reliability. Perhaps risk adjustment about extent of disability may also be useful.

**All well defined, no clarity gaps

**The developers report mean reliability scores of 0.939 for Medicare, 0.982 for Medicaid, and 0.935 for commercial, demonstrating that the measure is considered reliable.

**In the denominator description, it says: "The denominator includes individuals 18 years of age or older with one or more prescription claims for an opioid and a negative medication history for any opioid medication during the 90-day lookback period." I am not sure I understand the rationale of using a 90-day period. Why not 30 days, or 60 days?

**Specifications provided

**relaibility is there as it is claims based

2a2. Reliability – Testing
Comments:

**No

**No, claims based data should be fairly accurate and the entire population is sampled so selection bias should not be an issue

**No

**No

**No

**High – Reliability testing at the measure score level. Measured considered reliable based on the mean reliability scores of 0.939 for Medicare, 0.982 for Medicaid, and 0.935 for commercial.

**yes, co-factors ---- postop, etc are not fsctored

2b1. Validity – Testing

Comments:

**No

**As in number 6, making sure reasons for appropriate longer prescriptions are adequately adjusted.

**No

**No

**No

**Moderate – Independent 7 member Measure Validity Panel (MVP) agreed on the measure having face validity and can be used to distinguish high and low performing plans.

**Yes, not sure if a single refill means addiction

2b4-7. Threats to Validity

Comments:

**No

**The developers compared meaningful differences across medicare and medicaid population had significant variation but the commercial plans did not. A small number of patients were excluded due to exclusion criteria (unclear if this was validated, should there have been more exclude?) I did not see data on missing data, but expect it would be minimal and random

**No concerns.

**No

**No missing data were identified.

**No missing data found.

**There are factors like surgery, use of other medications not included

2b2-3. Other Threats to Validity

2b2. Exclusions

2b3. Risk Adjustment

Comments:

**Ok

**Would consider other exclusions - extent of diability? difficulty in accessing pharmacy? We would expect social risk factors to influence this metric. It may be valuable to stratify by social risk factors so as not to penalize providers or hospitals that serve high risk areas.

**Well considered and handled.

**Not applicable

**No risk adjustment was applied to the proposed measure.

**Exclusions appropriate. No risk adjustment.

**Why not use the state PMP data to see trends

Scientific Acceptability

Measure Number: 3558

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

Type of measure:

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

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

Data Source:

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

Level of Analysis:

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

Measure is:

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

RELIABILITY: SPECIFICATIONS

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

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

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

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

   No issues

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐ **No**

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

Submission document: Testing attachment, section 2a2.2

- Completed measure testing using administrative claims for the Medicare, Medicaid, and commercial populations.
- Data from Tester 1, including state Medicaid, were from January 1, 2017 – December 31, 2017.
- For Tester 2, the national Medicare data were from January 1, 2018 to December 31, 2018.
- Data from Tester 3, including state Medicaid and commercial, were from January 1, 2017 to December 31, 2017.
- A beta-binomial model was used to calculate plan-specific reliability scores.

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

Submission document: Testing attachment, section 2a2.3

- Based on the mean reliability scores of 0.939 for Medicare, 0.982 for Medicaid, and 0.935 for commercial, the measure is considered reliable

- **Table 4. Plan-Contract Reliability Score Distribution for 2018 National Medicare Data**

Statistic	Values
Mean	0.939
Standard Deviation	0.093
Interquartile Range	0.063
10th	0.788
25th	0.933
50th	0.984
75th	0.996
90th	0.999
Minimum	0.610
Maximum	0.999

- **Table 5. Plan Reliability Score Distribution for 2017 State Medicaid Sample**

Statistic	Values
Mean	0.982
Standard Deviation	0.023
Interquartile Range	0.015
10th	0.924
25th	0.980

50th	0.990
75th	0.995
90th	0.999
Minimum	0.924
Maximum	0.999

- **Table 6. Plan Reliability Score Distribution for 2017 Commercial Sample**

Statistic	Values
Mean	0.935
Standard Deviation	0.067
Interquartile Range	0.130
10th	0.861
25th	0.861
50th	0.953
75th	0.991
90th	0.991
Minimum	0.861
Maximum	0.991

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 the Reliability Algorithm:

Specifications precise unambiguous and complete (Box 1)→ Empirical reliability testing conducted (Box 2)→ Testing conducted at computed measure score level (Box 4)→ Method described and appropriate (Box 5) → Level of certainty or confidence that measure scores are reliable (Box 6) →HIGH (rationale that reliability improves as the sample sizes increase, medium and small facilities have lower reliability estimates)

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Exclusions:

- The developer supports exclusions (e.g., cancer, hospice, sickle cell) with evidence/guidelines from the CDC
- In testing, the sickle cell exclusion was only applied to the Medicare data, as it was identified as an appropriate exclusion after testing had concluded in the Medicaid and commercial data.
- The developer provided results to show the impact of exclusions:
- Tester 1: 0.35% (n=3,600) of patients were excluded based on receiving hospice and 0.33% (n=3,438) of patients were excluded based on a cancer diagnosis
- Tester 2: 3.61% (n=1,684,154) of patients were excluded based on a cancer diagnosis, 0.012% (n=5,813) of patients were excluded based on a sickle cell disease diagnosis, and 0.62% (n=288,075) of patients were excluded based on receiving hospice care.
- Tester 3: <.01% (n=50) of patients were excluded based on receiving hospice and 0.47% (n=7,302) of patients were excluded based on a cancer diagnosis.

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

Submission document: Testing attachment, section 2b4.

- 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.
- In accordance with the PQA measure specifications, measure rate calculations were only inclusive of plans with 30 or more individuals in the denominator.
- For the Medicare population, the measure rates showed significant variation, with an interquartile range of 13.5%. 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).
- For the Medicaid population, the measure rates showed significant variation, with an interquartile range of 11.9%. There is a statistically significant difference in measure rates between the top and bottom quartile of the plans included in the testing (P<0.01 at alpha = 0.05).
- For the commercial population, the measure rates did not show significant variation, with an interquartile range of 1.6%. A statistical test to determine the differences between the top and

bottom quartile of the three commercial plans was not included in the testing, as the data are not appropriate for this test.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
    Submission document: Testing attachment, section 2b5.
15. **Please describe any concerns you have regarding missing data.**
    Submission document: Testing attachment, section 2b6.
16. **Risk Adjustment**

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

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

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

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

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

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

16d.**Risk adjustment summary:**

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

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

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

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

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

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

VALIDITY: TESTING

17. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**
18. **Method of establishing validity of the measure score:**
    ☒ **Face validity**
    ☐ **Empirical validity testing of the measure score**
    ☒ **N/A (score-level testing not conducted)**
19. **Assess the method(s) for establishing validity**
    Submission document: Testing attachment, section 2b2.2
    - Conducted face validity testing of the computed performance score.
    - The developer convened an independent body, called the measure validity panel (MVP), which is made up of individuals not involved in the development or review of the measure concept or draft measure.
20. **Assess the results(s) for establishing validity**
    Submission document: Testing attachment, section 2b2.3

- The developer states that through discussion and vote, the MVP determines whether the performance measure scores have face validity.
- Of the 7 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.

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

    Submission document: Testing attachment, section 2b1.

    ☐ **Yes**

    ☐ **No**

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

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

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

    ☐ **Yes**

    ☐ **No**

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

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

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

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

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

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

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

    Guidance from Validity Algorithm:

    Specifications consistent with evidence (Box 1)→Potential threats to validity assessed (Box 2) → Empirical validity testing of measure as specified (Box 3)→ Face validity systematically assessed (Box 4) →Results indicate substantial agreement→MODERATE

ADDITIONAL RECOMMENDATIONS

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

    - None

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.

- The developer tested data elements collected by healthcare personnel during provision of care.
- Data elements are in defined fields in administrative claims.
- Use of this measure is subject to licensing agreements specified by the developer.

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?*
- *Is the data collection strategy ready to be put into operational use?*
- *Does the Committee have any feasibility concerns?*

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility

Comments:

**Highly feasible

**The data elements seem readily available

**Highly feasible. No e-concerns.

**No concerns - uses readily available claims data

**Claim data seem to be adequate.

**High – data captured in claims

**Very feasible

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

- The measure has been adapted for use in the CMS Enhanced Medication Therapy Management (EMTM) program.
- The developer indicates that the measure is planned to be added to the CMS Star Ratings program display page for 2023 and 2024. The developer indicates that CMS will consider adding the IOP-LD measure to the Star Ratings in the future pending rulemaking once we gain experience with the measure.

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

- This is a new measure so feedback is not yet available.
- The developer plans to utilize its consensus-based, multi-stakeholder measure development process to solicit feedback.

Additional Feedback:

- None, as this is a new measure.

***Questions for the Committee**:*

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

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

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

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

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

Improvement results

- This is a new measure so data on improvement is unavailable.
- The developer states that there is "opportunity for improvement, with mean measure rates among various lines of business ranging from 23.7% to 43.8%."

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

- This is a new measure, therefore, no unexpected findings were reported.
- The developer states that this measure is intended for use for retrospective population-level performance measurement and is not intended to guide clinical decision-making for individual patients

Potential harms

- This is a new measure, therefore no harm has been reported.

Additional Feedback:

- None

***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

Comments:

**Satisfies all requirements

**Part of the CMS EMTM public reporting which, it appears, will be available by public reporting

**Well dThis is a new measure so feedback is not yet availableescribed and adequate, aligns with other efforts outside NQF

**Sponsor stated that "certain uses of the measures may be subject to a licensing agreement specifying the terms of use and the licensing fee." Sponsor also indicated that PQA makes measure specifications available to its members before PQA endorsement votes". With these restrictions, it is not clear to me how the performance measure would be widely implemented at the national level.

**Used in CMS Enhanced Medication Therapy Management (EMTM) program. Planned for use in future CMS Star Ratings program.

**With new guidlines and strict documentation parameters, not sure what this will accomplish in addition

4b1. Usability – Improvement

Comments:

**Low risk of harms of under treatment

**The unintended consequence is undertreatment of pain. Coukld measurement of hospitalization or ED visit for pain be added as a measure to watch?

**New mesaure but concersn retained about unintended harms - less is not always better.

**The measure has been adapted for use in the CMS Enhanced Medication Therapy Management (EMTM) program, and the developer indicates that the measure is planned to be added to the CMS Star Ratings program display page for 2023 and 2024.

**No unintended consequences were identified. Still, it is not clear if and how the required licensing fee to use this measure would hamper a broad adaption of this measure.

**Moderate – Opportunity for improvement to narrow the gaps in performance by lines of business. Benefits outweigh harms.

**Moderate – Opportunity for improvement to narrow the gaps in performance by lines of business. Benefits outweigh harms.

Criterion 5: Related and Competing Measures

Related or competing measures

NQF 2940 : Use of Opioids at High Dosage in Persons Without Cancer
NQF 2950 : Use of Opioids from Multiple Providers in Persons Without Cancer
NQF 2951 : Use of Opioids from Multiple Providers and at High Dosage in Persons Without Cancer
NQF 3389 : Concurrent Use of Opioids and Benzodiazepines (COB)
NQF 3541 : Annual Monitoring for Persons on Long-Term Opioid Therapy (AMO)

Harmonization

- Measures are harmonized.
- Related measures each have different area of focus (numerator)
- There are no competing measures.

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

5. Related and Competing

Comments:

**There are numerous other metrics in this space

**No

**Yes - multiple - but all are aligned well and not duplicative

**None

**There are related existing measures on opioid prescribing, but they do not seem to compete with the proposed.

**5 related measures with different numerators. Harmonized.

**A few as mentioned

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 06/12/2020

- **Of the two NQF members who have submitted a support/non-support choice:**
  - One does not support the measure

Comment:

**The American Medical Association (AMA) strongly opposes the endorsement of NQF# 3558: Initial Opioid Prescribing for Long Duration (IOP-LD) as we believe that the measure is not aligned with the evidence as specified and there are significant unintended negative consequences that could be experienced with the use of this measure. The AMA believes that all care provided to patients must be individualized and quality measurement should not focus on preventing and/or reducing opioid use. Rather measurement should address the larger clinical issue—how well patients' pain is controlled, whether functional improvement goals are met, and what therapies are being used to manage pain while also lowering the risk of addiction and developing an opioid use disorder (OUD).

The ongoing singular focus on the dose and duration of opioid prescriptions disregards the important steps that the Administration has taken to address the national epidemic of opioid-related overdose deaths, which the AMA strongly supports. The final report of the Department of Health and Human Services (HHS) Interagency Pain Management Best Practices Task Force, for example, made a compelling case for the need to focus on patients experiencing pain as individuals and to develop treatment plans that meet their individual needs and not employ one-size-fits-all approaches that assume prescriptions of long duration are indications of overuse (HHS, 2019). Likewise, a Centers for Disease Control and Prevention (CDC) publication in the New England Journal of Medicine (Dowell, 2019) expressed concern that its opioid prescribing guidelines have been misapplied and wrongly used to discontinue or reduce prescriptions for patients with pain, with some actions likely to result in patient harm. The CDC stated that its guideline should not be used to create hard and fast policy; yet, it is the primary evidence provided to support this measure for accountability uses.

Specifically, the AMA does not believe that the evidence cited in support of the measure is sufficient since the CDC guidelines used the arbitrary 7-day threshold as a voluntary recommendation rather than a hard threshold. As the AMA warned in 2016, the CDC voluntary recommendation was taken beyond its context and used by state legislatures, pharmacy chains, pharmacy benefit managers, and health plans as authoritative to impose a hard, 7-day cap on opioid analgesic prescriptions and now we see it being used to hold a health plan accountable. Sole reliance on one guideline where the authors have explicitly voiced concerns with the inappropriate application of the recommendations should be avoided and we believe that the evidence subcriterion has not been met.

The AMA is further concerned that the measure uses a 90-day lookback period to define individuals who are “opioid naïve”. The CDC guideline does not define this population and the multiple studies cited throughout the measure submission form use varying timeframes (e.g., 60 days, 12 months). As a result, we believe that the use of a 90-day lookback period could drive inappropriate treatment decisions and the lack of an agreed upon definition for “opioid naïve” should prohibit this committee from determining that the measure as specified is evidence-based.

The AMA also believes that the numerator will incorrectly include those individuals who receive methadone for OUD treatment. Currently, the measure specifications consider methadone to be one of the opioid medications that should be included but because it does not exclude those patients with a diagnosis of OUD, anyone who receives one or more prescriptions for methadone for greater than 7 days will be considered to meet the numerator. We believe that the measure must address this error since it will lead to misrepresentations of performance and could lead to inappropriate treatment decisions in an effort to improve performance scores.

Lastly, the AMA is concerned with the usability of the measure and believes that there is significant potential for unintended negative consequences. While this measure is currently focused on health plan performance, there is great risk that it will lead todenials of medication in all instances even when an opioid is appropriately prescribed.

Given these significant concerns, the AMA does not support this measure and urges the Standing Committee not to recommend its endorsement.

References:

Dowell D, Haegerich T, Chou R. No shortcuts to safer opioid prescribing. N Engl J Med. 2019;380:2285–7. https://doi.org/10.1056/NEJMp1904190.

U.S. Department of Health and Human Services (2019, May). Pain Management Best Practices Inter-Agency Task Force Report: Updates, Gaps, Inconsistencies, and Recommendations. Retrieved from U. S. Department of Health and Human

Services website: https://www.hhs.gov/ash/advisory-committees/pain/reports/index.html.

- One supports the measure with clarification

Comment:

**The Federation of American Hospitals (FAH) appreciates the opportunity to comment on this measure prior to the Standing Committee's evaluation. The FAH recognizes the need to address inappropriate opioid use given the ongoing concerns around this important public health issue but we believe that measure must be aligned with evidence, provide useful information to accurately represent performance, and allow patients to make informed decisions.

The FAH requests that the committee consider whether the definition of "opioid naïve" used in this measure is aligned with current evidence and would not lead to inappropriate treatment decisions in an effort to improve performance scores. Specifically, the Centers for Disease Control and Prevention (CDC) guideline on which this measure is based does not explicitly define "opioid naïve" and the timeframes used in the other studies cited in the evidence form and throughout the submission vary from six months up to 12 months. As a result, it is not clear how the measure developers determined that a 90-day lookback period was the correct definition for "opioid naïve".

The FAH does not believe that measures used for accountability purposes should include specifications on which timeframes are selected in the absence of any consistent evidence and the resulting potential unintended negative consequences must be considered. The FAH requests that the committee discuss the lack of any evidence to support this lookback period and determine whether the measure as specified meets the NQF measure evaluation criteria.

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

IOP-LD_Evidence_Attachment_FINAL-637213515722530990.docx

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

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

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

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

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

Date of Submission: 4/2/2020

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

Outcome

☐ Outcome: Click here to name the health outcome

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

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

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

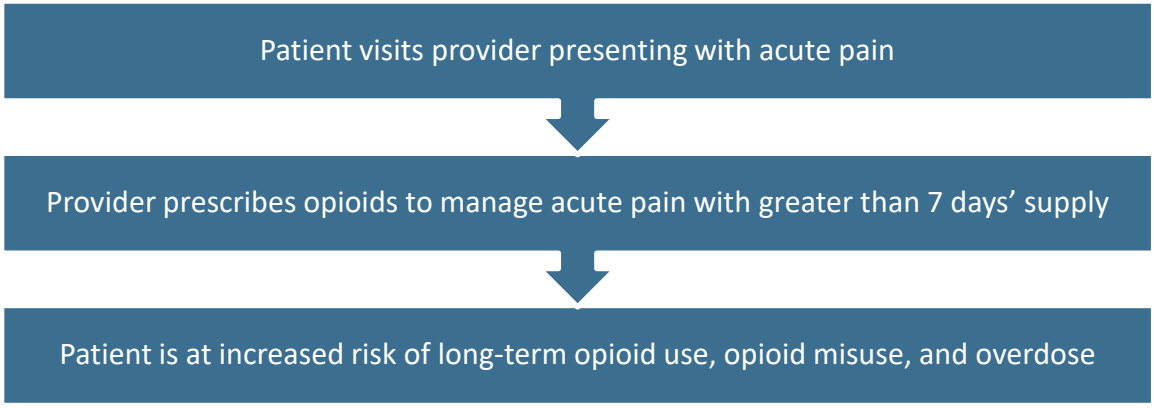
☒ Process: The percentage of individuals ≥18 years of age with ≥1 initial opioid prescriptions for >7 cumulative days' supply.

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

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

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



This measure evaluates the process of prescribing initial opioid therapy for a long duration (i.e., >7 days' supply). Clinical guidelines and a growing body of evidence indicate that when providers prescribe initial opioid therapy for a long duration, patients are at an increased risk of long-term opioid use, opioid misuse, overdose, and other adverse health outcomes. *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 when opioids are prescribed for the treatment of acute pain, providers should prescribe the lowest effective dose of immediate-release opioids and should prescribe no greater quantity than is needed for the expected duration of pain severe enough to require opioids.

This is aligned with a consistent body of evidence suggesting that a greater amount of early opioid exposure is associated with greater risk of long-term use, and by extension, a variety of adverse health outcomes such as fractures, endocrinologic harms, and other adverse health outcomes. According to the CDC Guideline's category A recommendation, three days' supply of opioids or less will often be sufficient, and more than seven days' supply will rarely be needed for the treatment of acute pain.

The CDC notes that physical dependence is an expected response in patients exposed to opioids for more than a few days. And Tehrani et al (cited in section 1a.4.1) suggest that given the more than 250 million prescriptions for opioid pain medications per year, even an average one day decrease in days' supply of opioid medications could have a beneficial effect on public health. By evaluating the percentage of individuals that receive initial opioid prescriptions for >7 days' supply, this performance measure drives this change by incentivizing health plans to identify and reduce initial opioid prescriptions that increase the risk of subsequent long-term opioid use, opioid misuse, overdose, and other adverse outcomes.

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

N/A

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

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

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

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

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

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 T, Chou R** • **March 18, 2016** • **Dowell D, Haegerich T, Chou R. CDC guideline for prescribing opioids for chronic pain — United States, 2016. Atlanta, GA: Centers for Disease Control and Prevention; 2016. https://www.cdc.gov/mmwr/volumes/65/rr/rr6501e1.htm. Accessed January 8, 2020.** • **https://www.cdc.gov/mmwr/volumes/65/rr/rr6501e1.htm**
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.	"Long-term opioid use often begins with treatment of acute pain. When opioids are used for acute pain, clinicians should prescribe the lowest effective dose of immediate-release opioids and should prescribe no greater quantity than needed for the expected duration of pain severe enough to require opioids. Three days or less will often be sufficient; more than seven days will rarely be needed." Recommendation Category: A, evidence type: 4
Grade assigned to the **evidence** associated with the	Evidence = Type 4: clinical experience and observations, observational studies with important limitations, or randomized clinical trials with several major limitations.

recommendation with the definition of the grade	Type 4 evidence indicates that one has very little confidence in the effect estimate, and the true effect is likely to be substantially different from the estimate of the effect.
Provide all other grades and definitions from the evidence grading system	Type 1 evidence: randomized clinical trials or overwhelming evidence from observational studies. Type 1 evidence: indicates that one can be very confident that the true effect lies close to that of the estimate of the effect. Type 2 evidence: randomized clinical trials with important limitations, or exceptionally strong evidence from observational studies. Type 2 evidence means that the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different. Type 3 evidence: observational studies or randomized clinical trials with notable limitations. Type 3 evidence: means that confidence in the effect estimate is limited and the true effect might be substantially different from the estimate of the effect.
Grade assigned to the **recommendation** with definition of the grade	Recommendation = Category A: apply to all persons in a specified group and indicate that most patients should receive the recommended course of action.
Provide all other grades and definitions from the recommendation grading system	Category B: indicate that there should be individual decision making; different choices will be appropriate for different patients, so clinicians must 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 – 2 studies Quality – Fair-quality retrospective cohort studies
Estimates of benefit and consistency across studies	Not applicable; the clinical evidence review did not identify benefits of initial opioid prescribing at long-duration.
What harms were identified?	The two studies identified in the clinical evidence review were consistent in their findings that opioid use for acute pain is associated with long-term opioid use, and that a greater amount of

	early opioid exposure is associated with greater risk for long-term use. The harms were estimated to be an increased for risk of opioid use at one year following initial opioid prescriptions within 7 days of surgery (adjusted OR 1.44, CI 1.39-1.50) and an increased risk for long-term use (defined as receiving five or more opioid prescriptions from 30-730 days) following opioid use within 15 days following onset of pain (adjusted OR 2.08, 95% CI 1.55 to 2.89 for 1-140 MME/day). Additionally, several guidelines on opioid prescribing for acute pain from emergency departments and other settings have recommended prescribing ≤3 days of opioids in most cases, whereas others have recommended ≤7 days or <14 days, and experts recommended a range of ≤3–5 days or ≤3–7 days when opioids are needed.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	1. Tehrani AB, Henke RM, Ali MM, Mutter R, Mark TL. Trends in average days' supply of opioid medications in Medicaid and commercial insurance. Addict Behav. 2017;76:218-222. PMID: 28858693. 2. Shah A, Hayes CJ, Martin BC. Characteristics of Initial Prescription Episodes and Likelihood of Long-Term Opioid Use – United States, 2006-2015. MMWR. 2017;66(10):265-269. PMID: 28301454. 3. Shah A, Hayes CJ, Martin BC. Factors Influencing Long-Term Opioid Use Among Opioid Naïve Patients: An Examination of Initial Prescription Characteristics and Pain Etiologies. J Pain. 2017;18(11):1374-1383. PMID: 28711636. 4. Zhang Y, Johnson P, Jeng P, Reid MC, Witkin L et al. First Opioid Prescription and Subsequent High-Risk Opioid Use: a National Study of Privately Insured and Medicare Advantage Adults. J Gen Intern Med. 2018;33(12):2156-62. PMID: 30206790. 5. Brat, G.A., Agniel, D., Beam, A., Yorkgitis, B., Bicket, M., Homer, M., Fox, K.P., Knecht, D.B., McMahill-Walraven, C.N., Palmer, N. and Kohane, I., 2018. Postsurgical prescriptions for opioid naive patients and association with overdose and misuse: retrospective cohort study. Bmj, 360, p.j5790. PMID: 29343479 6. Durand, Z., Nechuta, S., Krishnaswami, S., Hurwitz, E.L. and McPheeters, M., 2019. Prevalence and Risk Factors Associated With Long-term Opioid Use After Injury Among Previously Opioid-Free Workers. JAMA network open, 2(7), pp.e197222-e197222. PMID: 31314119. 7. Hadlandsmyth K, Lund BC, Mosher HJ. Associations between initial opioid exposure and the likelihood for long-term use. J Am Pharm Assoc . 2019;59(1):17-22. PMID: 30409501.

	The additional studies published since the literature review for the CDC Guideline provide additional empirical evidence to reinforce the Guideline's recommendation, and support the focus of the IOP-LD measure.

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.

Peer-reviewed literature published since the CDC's systematic review provides strong evidence to further support the CDC's recommendation that initial opioid prescriptions should not exceed seven days' supply.(1) This section synthesizes further evidence supporting the measure focus, with citations provided in chronological order.

A retrospective analysis by Tehrani et al. published in 2017 used Medicaid claims from 2005-2014, including approximately 6 million Medicaid enrollees ages 18-64 without a cancer diagnosis, to explore trends in prescribed days' supply for opioids.(2) The authors found that over the course of the study, the average days' supply of opioids increased considerably for all except for morphine. Oxycodone days' supply increased 4.5 days (a 37% increase) during the study period, while hydrocodone, hydromorphone, oxymorphone, and tapentadol increased by 4, 2.5, 3, and 5 days, respectively. Notably, there was no observed decline in median days' supplied for any opioid after 2013, when the opioid epidemic gained national attention. Further evaluation of a commercial cohort during the same timeframe exhibited similar, but even steeper increases in days' supply for the previously discussed opioids.

In research published in 2018, Shah et. al performed a retrospective cohort study using a random sample from a nationally representative database of the commercially insured population in the United States from 2006-2015 to explore relationships between initial opioid prescription characteristics and likelihood of long-term use.(3) The study examined 1,294,247 patients aged ≥18 years who met inclusion criteria and received initial opioid prescriptions, defined as those with no opioid prescriptions in the preceding 6 months of continuous enrollment. The study found that the probability of long-term opioid use increased with each additional day supplied when initiating opioid therapy, following the third day supplied. Specifically, the probability of long-term use was more than twice as high for individuals who received greater than 7 days' supply, when compared to those with at least one days' supply (13.5% vs. 6.0%). In conclusion, the authors expressly note that their findings are consistent with CDC recommendations and suggest that limiting initial opioid prescriptions to 7 or fewer days' supply (and ideally no greater than 3 days' supply) reduces the risk of unintentional long-term opioid use.

Additional research published by Shah et al. in 2018 provided further evidence that greater days' supply for initial opioid prescriptions is associated with a lower likelihood of opioid discontinuation.(4) In a retrospective cohort study examining a total of 1,353,902 opioid naïve individuals (defined as individuals with at least 12 months without an opioid prescription prior to their initial prescription) from 2006-2015 who met inclusion criteria, the authors examined the relationship between initial opioid prescription characteristics and probability of opioid discontinuation among opioid naïve patients, notably controlling for patient level factors. The authors reported that even controlling for patient factors and underlying pain etiologies, the results are consistent with earlier finding suggesting a dose-response relationship between days' supply and likelihood of discontinuation (see citation 3), with hazard ratios for discontinuation of 0.70 (95% CI 0.70-0.71) for a 3-4 day supply, 0.48 (95% CI 0.47-0.48) for a 5-7 day supply, 0.37 (95% CI 0.37-0.38) for an 8-10 day supply, and 0.32 for an 11-14 day supply of opioids (95% CI 0.31-33), where a 1-2 day supply is the reference group.

In 2018, Zhang et al. conducted a retrospective cohort study of 403,664 privately insured patients and 107,509 Medicare Advantage patients who initiated opioid therapy between 2011 and 2013, to determine how characteristics of these initial prescriptions, including days' supply, affect risk for high-risk opioid use in the following 18 months.(5) Initial opioid prescriptions were defined as opioid prescriptions among patients that did not have any opioid prescriptions within a lookback period of six months. The authors found that an initial opioid prescription consisting of greater than 7 days' supply, versus a 3 or fewer days' supply reference group, was associated with a significant ($p<0.001$) increase in high-risk opioid use, including overlapping opioid prescriptions (7% increase, 95% CI 6.2%-7%), concurrent use of opioids and benzodiazepines (8.7% increase, 95% CI 8.2%-9.2%), receiving opioids with a daily dosage of 120MME in the long term (4.8% increase, 95% CI 4.5%-5.2%), and use of opioids in each quarter of the 18 months following the index prescription (15% increase, 95% CI 15.0% -15.6%).

Brat et al. conducted a retrospective cohort study in 2018 that explored opioid prescribing patterns after surgery and found additional evidence regarding initial opioid prescribing and opioid misuse, defined as a composite of diagnosis codes for opioid abuse, dependence, or overdose.(6) The study used surgical claims from a linked medical and pharmacy administrative database from 2008-2016, and included 568,612 opioid naïve patients, defined as patients who had used 7 or fewer days of opioids in the 60 days preceding the surgery, and who received a postsurgical opioid. The authors note that duration of use, rather than dosage, was most strongly associated with opioid misuse, with an estimated increase rate of 20% of opioid misuse per each week of an opioid prescription after adjusting for covariates (95% CI 18.5%-21.4%). These findings remained consistent after sensitivity analyses, including removing the most common ICD code for opioid dependence and relying entirely on specific abuse and overdose codes.

A 2019 study by Durand et al. of 46,399 opioid-naive injured workers, defined as individuals with no opioid prescriptions in the 60 days prior to injury, from 2013 – 2015 examined the relationship between initial opioid prescription characteristics and long-term opioid use.(7) The authors found that initial prescription length, rather than demographic or injury characteristics, was the strongest predictor of long-term opioid use, which was defined as having an opioid supplied for 45 or more days in the 90 days after the injury, . The study found that in comparison to a reference group of <5 days' supply, a 5-9 day supply was associated with a significant increase in the odds of long-term use (adjusted OR 1.83, 95% CI 1.56-2.14). This dose-response trend between days' supply and odds of long-term opioid use continued with greater days' supply, including 10-19 days (adjusted OR 4.73, 95% CI 3.90-5.75) and 20 days or more (OR 28.94, 95% CI 23.44-35.72).

Finally, a 2019 study performed by Hadlandsmyth et al replicated Shah's methodology to examine the relationship between initial opioid exposures and long-term use. The study examined 19,600 patients in the Veteran's Health administration who received an initial opioid prescription (defined as prescriptions with no

opioid prescriptions in the preceding year) and met criteria for long-term opioid use within one year of follow-up. The authors corroborated Shah's findings, with initial opioid prescriptions for 7 days or fewer serving as a reference group, and greater days' supply associated with increased risk of long-term opioid use, including 8-14 days (adjusted OR 1.44, 95% CI 1.38-1.51), 15-21 days (adjusted OR 2.43, 95% CI 2.30-2.56), 22-30 days (adjusted OR 7.35, 95% CI 8.09-7.62), and greater than 30 days (adjusted OR 15.5, 95% CI 14.7-16.4).(8)

The studies synthesized in this section build on a consistent body of empirical evidence that reinforces the CDC's recommendation that initial opioid prescriptions should not exceed 7 days. The literature is fully consistent in finding that greater days' supply is associated with significant harms, including increased risk of long-term opioid use, opioid misuse, and overdose.

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 2020), and a search of the FDA website was conducted.

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

1. Dowell, D., Haegerich, T.M. and Chou, R., 2016. CDC guideline for prescribing opioids for chronic pain—United States, 2016. Jama, 315(15), pp.1624-1645. PMID: 26977696.
2. Tehrani AB, Henke RM, Ali MM, Mutter R, Mark TL. Trends in average days' supply of opioid medications in Medicaid and commercial insurance. Addict Behav. 2017;76:218-222. PMID: 28858693.
3. Shah A, Hayes CJ, Martin BC. Characteristics of Initial Prescription Episodes and Likelihood of Long-Term Opioid Use – United States, 2006-2015. MMWR. 2017;66(10):265-269. PMID: 28301454.
4. Shah A, Hayes CJ, Martin BC. Factors Influencing Long-Term Opioid Use Among Opioid Naïve Patients: An Examination of Initial Prescription Characteristics and Pain Etiologies. J Pain. 2017;18(11):1374-1383. PMID: 28711636.
5. Zhang Y, Johnson P, Jeng P, Reid MC, Witkin L et al. First Opioid Prescription and Subsequent High-Risk Opioid Use: a National Study of Privately Insured and Medicare Advantage Adults. J Gen Intern Med. 2018;33(12):2156-62. PMID: 30206790.
6. Brat, G.A., Agniel, D., Beam, A., Yorkgitis, B., Bicket, M., Homer, M., Fox, K.P., Knecht, D.B., McMahill-Walraven, C.N., Palmer, N. and Kohane, I., 2018. Postsurgical prescriptions for opioid naive patients and association with overdose and misuse: retrospective cohort study. Bmj, 360, p.j5790. PMID: 29343479.
7. Durand, Z., Nechuta, S., Krishnaswami, S., Hurwitz, E.L. and McPheeters, M., 2019. Prevalence and Risk Factors Associated With Long-term Opioid Use After Injury Among Previously Opioid-Free Workers. JAMA network open, 2(7), pp.e197222-e197222. PMID: 31314119.
8. Hadlandsmyth K, Lund BC, Mosher HJ. Associations between initial opioid exposure and the likelihood for long-term use. J Am Pharm Assoc . 2019;59(1):17-22. PMID: 30409501.

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.

Opioid misuse, addiction, and overdose are a public health crisis affecting social and economic welfare in the United States, with more than 130 Americans dying each day due to opioid overdose.(1) Although recent data suggest slight decreases in overdose deaths involving natural/semisynthetic opioids and methadone, overdose deaths involving synthetic opioids continue to grow, and the rates across all types of opioids remain unacceptably high.(2)

Prescription opioids for pain management remain a contributing factor to the crisis. A systematic review of 38 studies, conducted in 2015, found that approximately 21% to 29% of patients prescribed opioids for chronic pain misuse them, while the 2015 National Survey on Drug Use and Health found that 12.5% of adults with opioid prescriptions reported misuse, and 16.7% reported a prescription opioid use disorder.(3,4) In response to the opioid overdose epidemic, a public health emergency was declared in 2017 by the United States Department of Health and Human Services.(5)

The duration of initial opioid exposure is associated with a higher likelihood for high-risk and long-term opioid use, misuse, overdose, and other negative outcomes (6,7,8,9,10). The 2016 Centers for Disease Control and Prevention (CDC) Guideline for Prescribing Opioids for Chronic Pain recommends that when opioids are used for acute pain (i.e., pain with abrupt onset and caused by an injury or other process that is not ongoing), no greater quantity should be prescribed than is needed for the expected duration of pain severe enough to require opioids.(11) According to the guideline, three days or less will often be sufficient, and more than seven days of opioids will rarely be needed for the treatment of acute pain.

Published studies support the CDC recommendations to limit the duration of initial opioid use. In 2017, Shah and colleagues published a retrospective cohort study using claims data from a nationally representative database of commercially insured patients, examining the relationship between initial opioid prescription characteristics and the likelihood of opioid discontinuation.(6) Increasing days' supply of the first prescription was consistently associated with a lower likelihood of opioid discontinuation: 3-4 days' supply (Hazard Ratio [HR], 0.70; 95% Confidence Interval [CI], 0.70-0.71); 5-7 days' supply (HR, 0.48; 95% CI, 0.47-0.48); 8-10 days' supply (HR, 0.37; 95% CI, 0.37-0.38); 11-14 days' supply (HR, 0.32; 95% CI, 0.31-0.33).

Zhang and colleagues found that an initial opioid prescription of greater than seven days' supply was associated with a significant ($p<0.001$) increase in high-risk opioid use, including overlapping opioid prescriptions, concurrent use of opioids and benzodiazepines, and the probability of three or more opioid prescribers, as well as increased likelihood of long-term opioid use.(7) In addition, Brat and colleagues published a study in 2018 evaluating the effects of varying opioid prescribing patterns after surgery on misuse or overdose in a retrospective cohort study of an opioid naïve population. (8) The total duration of opioid use was the strongest predictor of misuse, with an estimated increase rate of 20% of opioid misuse per each week of an opioid prescription after adjusting for covariates. A study by Durand et al of 46,399 opioid-naive injured workers found in comparison to a reference group of <5 days' supply, an initial prescription with 5-9 days' supply was associated with a significant increase in the odds of long-term use (adjusted OR 1.83, 95% CI 1.56-2.14).(9) Hadlandsmyth and colleagues found that days' supply was the strongest predictor in a multivariable model of long-term opioid use, with initial opioid prescriptions for 7 days or fewer serving as a reference group, and greater days' supply associated with increased risk of long-term opioid use, including 8-14 days

(adjusted OR 1.44, 95% CI 1.38-1.51).(10) For a more detailed review of the evidence, please refer to the evidence form.

Aligned with CDC recommendations and published evidence, this performance measure evaluates the percentage of individuals 18 years of age and older with one or more initial opioid prescriptions for >7 cumulative days' supply. Patients with cancer diagnoses, sickle cell diagnoses, and those receiving hospice care are excluded from the measure because of their unique therapeutic goals, ethical considerations, opportunities for medical supervision, and balance of risks and benefits.(11, 12) This measure was designed for retrospectively evaluating health plan performance at the population level and is not intended to guide clinical care for individual patients.

Citations:

1) NIH/NIDA. Opioid Overdose Crisis [Internet]. 2019 [cited 2019 Mar 29]. Available from: https://www.drugabuse.gov/drugs-abuse/opioids/opioid-overdose-crisis.

2) Hedegaard H, Minino AM. Warner M Drug Overdose Deaths in the United States, 1999-2018. NCHS Data Brief No 365. January 2020. Available from: https://www.cdc.gov/nchs/products/databriefs/db356.htm

3) Han B, Compton WM, Blanco C, et al. Prescription Opioid Use, Misuse, and Use Disorders in U.S. Adults: 2015 National Survey on Drug Use and Health. Ann Intern Med. 2017;167(5):293-301. PMID: 28761945.

4) Vowles KE, McEntee ML, Julnes PS, Frohe T, Ney JP, van der Goes DN. Rates of opioid misuse, abuse, and addiction in chronic pain: a systematic review and data synthesis. Pain. 2015; 156:569-76. PMID: 25785523.

5) HHS. National Rx Drug Abuse and Heroin Summit. Secretary Price Announces HHS Strategy for Fight Opioid Crisis. 2017. Available from: https://www.hhs.gov/about/leadership/secretary/speeches/2017-speeches/secretary-price-announces-hhs-strategy-for-fighting-opioid-crisis/index.html.

6) Shah A, Hayes CJ, Martin BC. Factors Influencing Long-Term Opioid Use Among Opioid Naive Patients: An Examination of Initial Prescription Characteristics and Pain Etiologies. J Pain. 2017; 18:1374-83. PMID: 28711636.

7) Zhang Y, Johnson P, Jeng PJ, et al. First Opioid Prescription and Subsequent High-Risk Opioid Use: a National Study of Privately Insured and Medicare Advantage Adults. J Gen Intern Med. 2018;33(12):2156-2162. PMID: 30206790.

8) Brat GA, Agniel D, Beam A, et al. Postsurgical prescriptions for opioid naive patients and association with overdose and misuse: retrospective cohort study. BMJ. 2018; 360:j5790. PMID: 29343479.

9) Durand, Z., Nechuta, S., Krishnaswami, S., Hurwitz, E.L. and McPheeters, M., 2019. Prevalence and Risk Factors Associated With Long-term Opioid Use After Injury Among Previously Opioid-Free Workers. JAMA network open, 2(7), pp.e197222-e197222. PMID: 31314119.

10) Hadlandsmyth K, Lund BC, Mosher HJ. Associations between initial opioid exposure and the likelihood for long-term use. J Am Pharm Assoc . 2019;59(1):17-22. PMID: 30409501.

11) Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain - United States, 2016. MMWR Recomm Rep. 2016;65:1-49. PMID: 26987082.

12) Carlon RW, Hudis CA, Ligget M. 2019 CDC Opioid Guideline Clarification Letter to ASCO / ASH / NCCN. 2019 Feb. Centers for Disease Control and Prevention. Available from 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.

1b.2. Provide performance scores on the measure as specified (<u>current and over time</u>) at the specified level of analysis. *(<u>This is required for maintenance of endorsement</u>. 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 Pharmacy Quality Alliance (PQA) completed measure testing using administrative claims data for Medicare, Medicaid, and commercial populations. The data sources included state Medicaid, national Medicare, and an assortment of commercial plans.

State Medicaid data were from January 1, 2017 - December 31, 2017 and included nine health plans with a total of 728,645 individuals aged 18 and older. National Medicare data were from January 1, 2018 to December 31, 2018 and included 673 Medicare Advantage Prescription Drug (MAPD) plans and 61 standalone Prescription Drug Plans (PDPs) with a total of 46,675,451 individuals aged 18 and older. Commercial data were from January 1, 2017 to December 31, 2017 and included a total of three health plans across three states with a total of 1,266,256 individuals aged 18 and older. Data from these time periods were the most recent, complete, full-year data available to testers at the time of testing.

Plan/contract-level Medicare rates ranged from 16.7% to 86.6%, with a mean rate of 43.8%, a median rate of 41.8%, and a standard deviation of 11.9%. A Student's t-test found statistically significant differences between performance in the 25th and 75th percentiles.

Plan/contract-level Medicaid rates ranged from 9.5% to 33.5%, with a mean of 23.7%, a median rate of 25.9%, and a standard deviation of 8.1%. A Student's t-test found statistically significant differences between performance in the 25th and 75th percentiles.

Plan/contract-level commercial rates ranged from 23.7% to 26.8%, with a mean of 25.1%, a median of 24.7%, and a standard deviation of 1.6%. A Student's t-test was not performed, as there were too few plans in the commercial sample to determine significant differences between the 25th and 75th percentiles.

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.

As described in 1b.2, our testing revealed significant opportunity for improvement. Our findings are consistent with the existing body of evidence, which highlights initial opioid prescription as an area with a performance gap. For example, a CDC Morbidity and Mortality Weekly Report found that in a sample of 1,292,247 patients enrolled in commercial health plans from 2006-2015, approximately 30% of opioid naïve patients had an initial duration of opioids of >7 days, and 7.3% were initially prescribed opioids for greater than or equal to 31 days.(1) Additionally, Mundkur et al. found that in a sample of 205,560 individuals from 2014, 46% of initial opioid prescriptions supplied were greater than a seven days' supply, and 10% of prescriptions supplied were for 30 days' supply or greater.(2) Among a sample of over ten million commercially insured individuals from 2012-2017 who had not used opioids in the prior six months and received their first opioid prescription during the study period, Zhu et al. found that approximately 57% of initial prescriptions were for more than a 3-day supply, and nearly 16% exceeded a 7-day supply.(3)

1) Shah A, Hayes CJ, Martin BC. Characteristics of Initial Prescription Episodes and Likelihood of Long-Term Opioid Use - United States, 2006-2015. MMWR Morb Mortal Wkly Rep. 2017;66(10):265-269. Published 2017 Mar 17. PMID: 28301454.

2) Mundkur, ML, Rough, K, Huybrechts, KF, et al. Patterns of opioid initiation at first visits for pain in United States primary care settings. Pharmacoepidemiol Drug Saf. 2018; 27: 495- 503. PMID: 28971545.

3) Zhu W, Chernew EM, Sherry TB, Maesyas N. Initial Opioid Prescriptions among U.S. Commercially Insured Patients, 2012-2017. N Engl J Med 2019; 380:1043-1052. PMID: 30865798.

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.*

For all lines of business, measure rates were available and analyzed based on age and gender using the samples described in 1b.2.

For Medicare, the beneficiary-level Low Income Subsidy (LIS) variable was used to determine differences 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 Medicaid and commercial, LIS is not reported, and no other patient-level indicators of sociodemographic status were available in the data.

In the Medicare data, measure rates were 36.4% among males and 43.6% among females. For age bands, measure rates were 30.2% in 18-50, 45.1% in 51-64, 39.5% in 65-85, and 52.7% in 85+. Measure rates were 44.7% among LIS, and 39.1% among non-LIS.

In the Medicaid data, measure rates were 19.9% among males and 14.7% among females. For age bands, measure rates were 11.5% in 18-50, 30.8% in 51-64, 52.0% in 65-84, and 58.1% in 85+.

In the commercial data, measure rates were 26.9% among males and 25.8% among females. For age bands, measure rates were 21.5% in 18-50, 34.2% in 51-64, 36.5% in 65-84, and 58.8% among 85+.

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

For most studies on initial opioid prescribing cited in the evidence form and in the measure rationale, data on race/ethnicity, socioeconomic status, and other demographic variables were unavailable or not discussed. However, Durand et al, whose study is described further in 1b.1 and the evidence form, found that in their model, higher odds of long-term opioid use were associated with rural residence compared with urban residence (OR, 1.51; 95% CI 1.31-1.73). (1)

Other evidence suggests the presence of racial and urban/rural disparities in adverse opioid-related outcomes such as overdose. The CDC has found that whites and American Indian or Alaska Natives are more likely to overdose on prescription painkillers: in 2008, the rate of deaths involving opioids were roughly three times higher among these racial groups versus blacks and Hispanic whites. (2, 3) Additionally, the CDC has found that individuals in rural counties are more likely to overdose on prescription painkillers, versus urban settings, with underlying research from Paulozzi et al suggesting age-adjusted rates of 3.85 and 2.76 opioid poisoning fatalities per 100,000, respectively. (2, 4)

Recent research by Altekruse et al used the Mortality Disparities in American Community Study, which links nearly 4 million 2008 American Community Survey responses to the National Death Index from 2008-2015, to shed further light on disparities in opioid overdose deaths. (5) Their findings on race and ethnicity were consistent with CDC, as whites (HR 2.52, 95% CI 2.21-2.88) and American Indians/Alaska Natives (HR 1.88, 95% CI 1.35-2.62) had higher risk for overdose versus Hispanics. They also examined various socioeconomic factors, finding elevated risk for individuals living in poverty versus those five times or higher above the poverty line (HR 1.36, 95% CI 1.20-1.54), and higher risk for people without health insurance versus those with health insurance (HR 1.30, 95% CI 1.20-1.41). The study found elevated risk among men (HR 1.61, 95% CI 1.50-1.72) and among those who are disabled (HR 2.80, 95% CI 2.59-3.03). The study's findings differed from other research on rural/urban disparities, finding higher risk for non-rural residents versus rural residents (HR 1.46, 95% CI 1.34-1.59).

Citations:

1) Durand, Z., Nechuta, S., Krishnaswami, S., Hurwitz, E.L. and McPheeters, M., 2019. Prevalence and Risk Factors Associated With Long-term Opioid Use After Injury Among Previously Opioid-Free Workers. JAMA network open, 2(7), pp.e197222-e197222. PMID: 31314119.

2) Vital Signs: Prescription Painkiller Overdoses in the US. N.d. Centers for Disease Control and Prevention. Available from https://www.cdc.gov/vitalsigns/PainkillerOverdoses/index.html#science.

3) Centers for Disease Control and Prevention (CDC). Vital signs: overdoses of prescription opioid pain relievers--United States, 1999--2008. MMWR Morb Mortal Wkly Rep. 2011;60(43):1487–1492. PMID: PMID: 22048730.

4) Paulozzi LJ, Xi Y. Recent changes in drug poisoning mortality in the United States by urban-rural status and by drug type. Pharmacoepidemiol Drug Saf. 2008;17(10):997–1005. PMID: 18512264.

5) Altekruse SF, Cosgrove CM, Altekruse WC, Jenkins RA, Blanco C. Socioeconomic risk factors for fatal opioid overdoses in the United States: Findings from the Mortality Disparities in American Communities Study (MDAC). PLoS One. 2020;15(1):e0227966. Published 2020 Jan 17. PMID: 31951640.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

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

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

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

Not applicable.

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

This is not an eMeasure **Attachment:**

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

Attachment **Attachment:** PQA_IOP_Value_Sets-637124369595574869.xlsx

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

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

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

Not an instrument-based measure

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

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

Not applicable.

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

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

The number of individuals from the denominator with >7 cumulative days' supply for all opioid prescription claims within any opioid initiation period.

The opioid initiation period is defined as the date of service of the initial opioid prescription plus two days, i.e., the 3-day time period when the numerator is assessed.

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 >7 cumulative days' supply for all opioid prescription claims within any opioid initiation period.

Use the steps below to identify the numerator population:

Step 1: For each individual in the denominator population, identify all initial opioid prescriptions and corresponding opioid initiation periods, defined as the date of service of the initial opioid prescription plus two days.

For example, if the date of service for an initial opioid prescription is March 15, identify all opioid prescription claims from March 15 through March 17.

Step 2: For each individual, starting with each initial opioid prescription, sum the days' supply of all opioid prescriptions within each opioid initiation period.

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 claim 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.

• If the opioid initiation period extends beyond the end of the measurement year, the opioid initiation period is truncated to the last day of the measurement year.

Step 3: Count the unique individuals with >7 cumulative days' supply for all opioid prescription claims during any opioid initiation period in the measurement year.

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

The denominator includes individuals 18 years of age or older with one or more prescription claims for an opioid and a negative medication history for any opioid medication during the 90-day lookback period.

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

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

The denominator includes individuals aged 18 years or older as of the first day of the measurement year with at least one prescription claim for an opioid medication during the measurement year, with continuous enrollment during the measurement year and 90 days prior to the index prescription start date (IPSD) and a negative medication history for any opioid medication during the 90-day lookback period.

Individuals in hospice at any time during the measurement year or 90 days prior to the first day of the measurement year, and those with a cancer or sickle cell disease diagnosis during the measurement year or 90 days prior to the first day of the measurement year are excluded from the measure.

Complete the steps below to determine the denominator population.

Step 1: Identify individuals 18 years or older as of the first day of the measurement year.

Step 2: Identify individuals with one or more prescription claims for an opioid (Medication Table OPIOIDS) during the measurement year.

Step 3: Identify individuals continuously enrolled during the measurement year and the 90 days prior to the IPSD.

Step 4: Identify unique individuals with a negative medication history for any opioid medication during the 90-day lookback period.

For example, an individual has opioid prescription claims on August 1, September 15 and December 20. For each of these dates of service, use the lookback period of 90 days to determine if the individual had no prescription claims for opioids (Medication Table OPIOIDS). For example, for August 1, determine whether the individual had no prescription claims for opioids from May 3 - July 31. Repeat for the September 15 and December 20 opioid prescription claims.

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 claim 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.

• Count the unique individuals (i.e., if an individual has multiple lookback periods, count the individual only once in the denominator).

Step 5: Exclude individuals with any of the following during the measurement year or the 90 days prior to the first day of the measurement year:

• Hospice

• Cancer

• Sickle Cell Disease

Medication Table OPIOIDS: Opioids

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

(Note: Includes combination products. Excludes the following: injectable formulations; opioid cough and cold products; sublingual sufentanil [used in a supervised setting]; and all buprenorphine products, as buprenorphine, as a partial opioid agonist, is not expected to be associated with overdose risk in the same dose-dependent manner as doses for full agonist opioids.)

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 or the 90 days prior to the first day of 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 individuals in hospice during the measurement year or 90 days prior to the first day of the measurement year. To identify individuals in hospice:

•Hospice indicator from the enrollment database, if available (e.g. Medicare)

•One or more claims with place of service code 34 during the measurement year or 90 days prior to the first day of the measurement year, if hospice indicator is not available (e.g. Commercial, Medicaid)

Cancer exclusion: Exclude any individuals with cancer during the measurement year or 90 days prior to the first day of the measurement year.

•One or more claims with cancer in the primary diagnosis or any other diagnosis fields during the measurement year or 90 days prior to the first day of the measurement year. See PQA ICD Code Value Sets, Cancer tab.

•Pharmacy hierarchical condition category (RxHCC) 15, 16, 17, 18, 19 from the Medicare Part D risk adjustment model for payment year 2017 or 2018, if ICD codes are not available. [Available from https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.html]

Sickle cell exclusion: Exclude any individuals having one or more claims with sickle cell disease (SCD) in the primary diagnosis or any other diagnosis fields during the measurement year or 90 days prior to the first day of the measurement year. See PQA ICD Code Value Sets, SickleCellDisease tab.

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

The measure is stratified by the following lines of business for the health plan:

•Commercial

•Medicare

•Medicaid

Medicare plans are further stratified by Low Income Subsidy status.

Definition: Medicare Low Income Subsidy (LIS) - 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. Medicare beneficiaries apply for the LIS with the Social Security Administration or their State Medicaid agency.

The Medicare Master Beneficiary Summary file contains the Cost Share Group variable used to identify Low Income Subsidy status, which is subsidized Part D coverage. There are 12 monthly variables - where the 01 through 12 at the end of the variable name corresponds with the month (e.g., 01 is January and 12 is December). CMS identifies beneficiaries with fully subsidized Part D coverage by looking for individuals that have a 01, 02, or 03 for the month. Other beneficiaries who are eligible for the LIS but do not receive a full subsidy have a 04, 05, 06, 07, or 08. The remaining values indicate that the individual is not eligible for subsidized Part D coverage.

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 or older as of the first day of the measurement year.

Step 2: Identify individuals with one or more prescription claims for an opioid (see Medication Table OPIOIDS, below) during the measurement year.

Step 3: Identify individuals continuously enrolled during the measurement year and the 90 days prior to the IPSD.

Step 4: Identify unique individuals with a negative medication history for any opioid medication during the 90-day lookback period.

For example, an individual has opioid prescription claims on August 1, September 15 and December 20. For each of these dates of service, use the lookback period of 90 days to determine if the individual had no prescription claims for opioids. For example, for August 1, determine whether the individual had no prescription claims for opioids from May 3 - July 31. Repeat for the September 15 and December 20 opioid prescription claims.

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 claim 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.

• Count the unique individuals (i.e., if an individual has multiple lookback periods, count the individual only once in the denominator).

Step 5: (Exclusions) Identify individuals with any of the following during the measurement year or the 90 days prior to the first day of the measurement year:

- Hospice: Individuals in hospice during the measurement year or 90 days prior to the first day of the measurement year. Identify individuals in hospice using:
  - Hospice indicator from the enrollment database, if available (e.g. Medicare); or
  - One or more claims with place of service code 34 during the measurement year or 90 days prior to the first day of the measurement year, if hospice indicator is not available (e.g. Commercial, Medicaid)
- Cancer: Identify individuals with cancer during the measurement year or 90 days prior to the first day of the measurement year. Identify individuals with cancer using:
  - One or more claims with cancer in the primary diagnosis or any other diagnosis fields during the measurement year or 90 days prior to the first day of the measurement year. See PQA ICD Code Value Sets, Cancer tab.

o Pharmacy hierarchical condition category (RxHCC) 15, 16, 17, 18, 19 from the Medicare Part D risk adjustment model for payment year 2017 or 2018, if ICD codes are not available. [Available from https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.html]

- Sickle Cell Disease: Identify individuals having one or more claims with sickle cell disease (SCD) in the primary diagnosis or any other diagnosis fields during the measurement year or 90 days prior to the first day of the measurement year. See PQA ICD Code Value Sets, SickleCellDisease tab.

Table OPIOIDS: Opioids

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

(Note: Includes combination products; Excludes the following: injectable formulations; opioid cough and cold products; sublingual sufentanil [used in a supervised setting]; and all buprenorphine products, as buprenorphine, as a partial opioid agonist, is not expected to be associated with overdose risk in the same dose-dependent manner as doses for full agonist opioids.)

Step 6: Subtract the individuals identified in Step 5 (exclusions) from the population identified through Steps 1-4. The remaining individuals represent the denominator.

B. Numerator Population:

Step 7: For each individual in the denominator population, identify all initial opioid prescriptions and corresponding opioid initiation periods.

Step 8: For each individual, starting with each initial opioid prescription, sum the days' supply of all opioid prescriptions within each opioid initiation period (i.e., the initial opioid prescription + 2 days).

For example, if the date of service for an initial opioid prescription is March 15, identify any opioid prescription claims from March 15 through March 17.

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 claim 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.
- If the opioid initiation period extends beyond the end of the measurement year, the opioid initiation period is truncated to the last day of the measurement year.

Step 9: Count the unique individuals with >7 cumulative days' supply for all opioid prescription claims during any opioid initiation period in the measurement year. 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.

- Note: Report the rates separately by line of business (e.g. Medicare, Medicaid, Commercial). For Medicare, report rates for low-income subsidy (LIS) and non-LIS populations separately.

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

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

Not applicable.

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

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

Not applicable.

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, Prescription Drug Hierarchical Condition Categories (RxHCCs); 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.)*

Not applicable because this is not a composite performance measure.

2. Validity – See attached Measure Testing Submission Form

IOP-LD_Testing_Form_FV_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.

2.2 For maintenance of endorsement

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

2.3 For maintenance of endorsement

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

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

Measure Number (*if previously endorsed*)**:** N/A
Measure Title: Initial Opioid Prescribing for Long Duration (IOP-LD)
Date of Submission: 1/6/2020

Type of Measure:

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

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

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

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

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

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 Pharmacy Quality Alliance (PQA) completed measure testing using administrative claims data for the Medicare, Medicaid, and commercial populations. The measure was tested in full by three testers. The data sources from these three testers were from state Medicaid, national Medicare, and an assortment of commercial plans.

Tester	Organization	Populations
Tester 1	State Medicaid Agency	State Medicaid (MCO and FFS)
Tester 2	Federal Contractor	National Medicare (MAPD and PDP)
Tester 3	Analytic Firm	State Medicaid (MCO), Commercial

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

Data from Tester 1, including state Medicaid, were from January 1, 2017 – December 31, 2017. For Tester 2, the national Medicare data were from January 1, 2018 to December 31, 2018. Data from Tester 3, including state Medicaid and commercial, were from January 1, 2017 to December 31, 2017. Data from these time periods were the most recent, complete, full year data available to testers at the time of testing.

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

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 population was tested using the national 100% Medicare data set. The eligible population for the measure (i.e., the denominator) includes: individuals aged 18 years or older by the first day of the measurement year with at least one prescription claim for opioid medications during the measurement year, with continuous enrollment during the measurement year and 90 days prior to the index prescription start date (IPSD) and a negative medication history for any opioid medication during the 90-day lookback period. Individuals in hospice at any time during the measurement year or 90 days prior to the first day of the measurement year, and those with a cancer or sickle cell disease diagnosis during the measurement year or 90 days prior to the first day of the measurement year are excluded from the measure.

After applying the eligible population criteria, the resulting data included 673 Medicare Advantage Prescription Drug (MAPD) plans and 61 standalone Prescription Drug Plans (PDPs). Plans varied in size (see Table 1), with a mean plan size of 9,354 beneficiaries and a median plan size of 767 beneficiaries.

Table 1. Plan Size Distribution for 2018 National Medicare After Applying Eligible Population Criteria

Statistic	Number of Beneficiaries
Mean	9,354
Standard Deviation	53,741
Minimum	0
25th Percentile	78
50th Percentile	767
75th Percentile	3,754
Maximum	969,949
Interquartile Range	3,676

The Medicaid population was tested using state Medicaid data from two testers. After applying the eligible criteria population, the resulting data included 9 health plans across four states: UT (5), TN (2), PA (1), WV (1). Notably, the two MCOs in TN were larger than the plans in the other states. The Medicaid testing did not include the sickle cell disease exclusion. However, as discussed in section 2b.2, we found this exclusion to be rare and have minor effects on measure rates. Of the 9 plans, 1 plan was fee-for-service (FFS), and the remaining 8 plans were Medicaid Managed Care plans (MCOs). Plans varied in size (see Table 2), with a mean plan size of 9,402 beneficiaries and a median plan size of 2,297 beneficiaries.

Table 2. Plan Size Distribution for 2017 State Medicaid After Applying Eligible Population Criteria

Statistic	Number of Beneficiaries
Mean	9,402
Standard Deviation	14,846
Minimum	405
25th Percentile	1,473
50th Percentile	2,297
75th Percentile	6,239
Maximum	44,259
Interquartile Range	4,766

The commercial population was tested using private health plan data. After applying the eligible criteria population, the resulting data included three health plans across three states. Notably, the commercial testing did not include the sickle cell disease exclusion. However, as discussed in section 2b.2, we found this exclusion to be rare and have minor effects on measure rates. Plans varied in size, with a median plan size of 21,108 beneficiaries, maximum plan size of 119,878, and minimum plan size of 6,615. Because there are only three plans, further descriptive statistics are not provided.

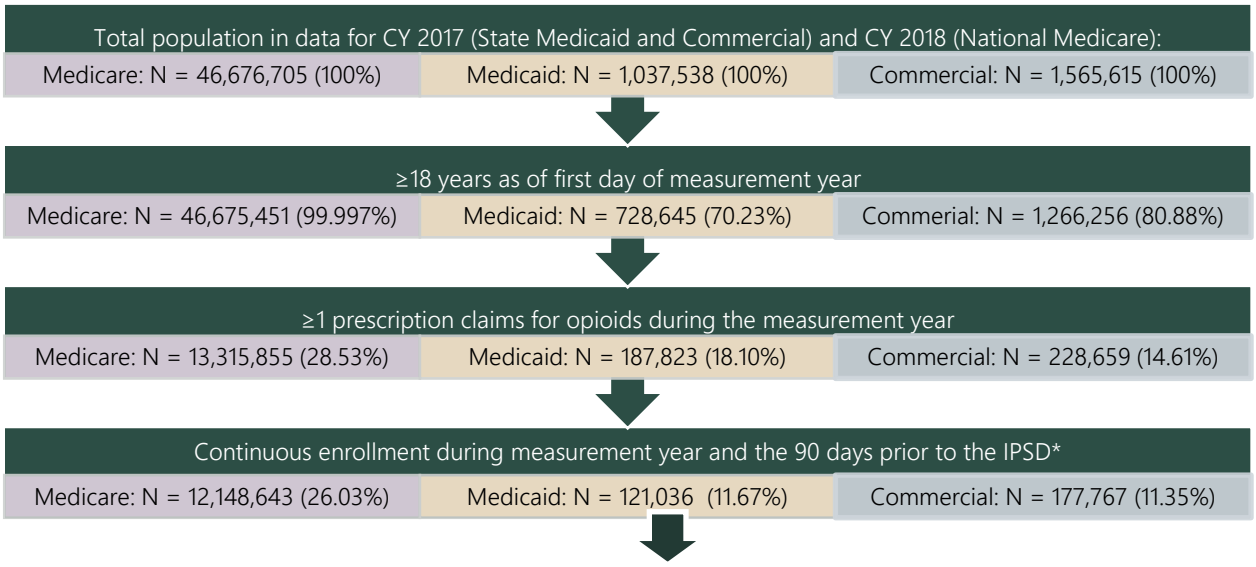
1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected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

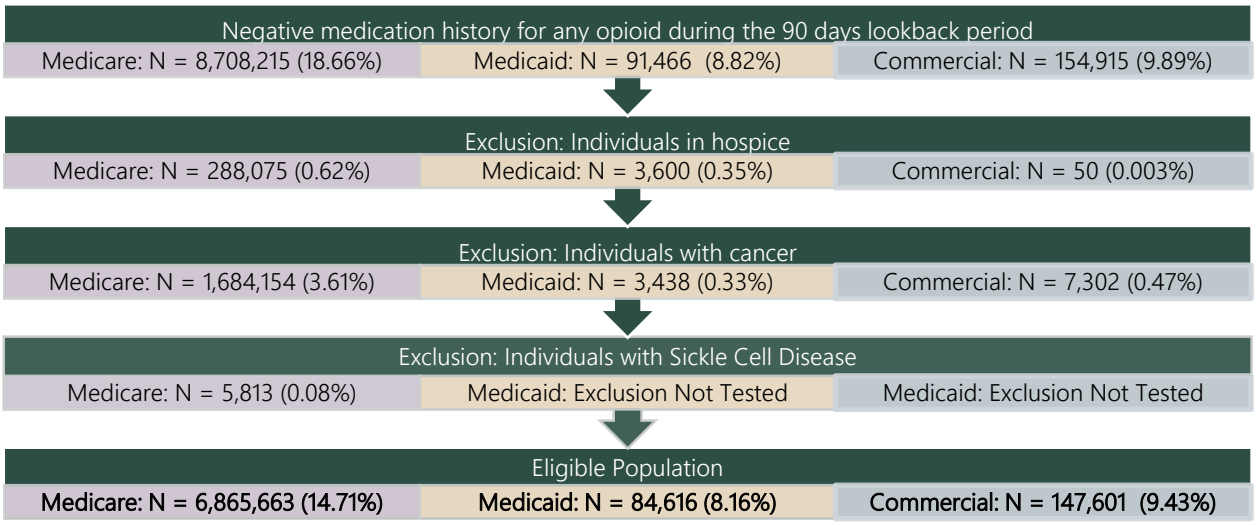
For the Medicare testing, a total of 46,675,451 individuals aged 18 and older were included in the testing and analysis. After applying all inclusion and exclusion criteria, the eligible population for analysis was 6,865,663 (14.72%) of the initial population. Of eligible persons, 2,707,353 (39.43%) were male. Individuals by age group included 420,956 (6.13%) age 18 – 50 years, 863,656 (12.58%) age 51 – 64 years, 5,001,559 (72.85%) age 65 – 84 years, and 579,592 (8.44%) age 85 and older.

For the Medicaid testing, a total of 728,645 individuals aged 18 and older were included in the testing and analysis. After applying all inclusion and exclusion criteria, the eligible population for analysis was 84,616 (8.16%) of the initial population. Of eligible persons, 21,820 (25.79%) were male. Individuals by age group included 66,754 (78.89%) age 18-50 years, 16,121 (19.05%) age 51-64 years, 1,648 (1.95%) age 65-84 years, and 93 (0.11%) age 85 and older.

For the commercial testing, a total of 1,266,256 individuals aged 18 and older were included in the testing and analysis. After applying all inclusion and exclusion criteria, the eligible population for analysis was 147,601 (9.43%) of the initial population. Of eligible persons, 61,062 (41.37%) were male. Individuals by age group included 92,348 (62.57%) age 18-50 years, 52,915 (35.85%) age 51-64 years, 2,304 (1.56%) age 65-84 years, and 34 (.02%) age 85 and older.

Table 3. Eligible Population by Criterion





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

The only difference to highlight was that the state Medicaid testers (Testers 1 and 3) and the commercial tester (Tester 3) did not test for, and therefore did not exclude, individuals with sickle cell disease. The national Medicare data testing did exclude individuals with sickle cell disease. This was not due to any data limitations, but rather because sickle cell disease was identified and added as an exclusion after testing had concluded in the Medicaid and commercial populations.

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 all lines of business, measure rates were available and analyzed based on age and gender.

For Medicare, the beneficiary-level Low Income Subsidy (LIS) variable was used to determine differences 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 Medicaid, LIS is not reported, and no other 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 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. In accordance with the PQA measure specifications, reliability testing excluded plans with less than 30 individuals in the denominator.

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 4 shows the distribution of the plan-contract-level scores for Medicare, Table 5 shows the plan-level scores for Medicaid, and Table 6 shows the plan-level scores for commercial.

Table 4. Plan-Contract Reliability Score Distribution for 2018 National Medicare Data

Statistic	Values
Mean	0.939
Standard Deviation	0.093
Interquartile Range	0.063
10th	0.788
25th	0.933
50th	0.984
75th	0.996

90th	0.999
Minimum	0.610
Maximum	0.999

Table 5. Plan Reliability Score Distribution for 2017 State Medicaid Sample

Statistic	Values
Mean	0.982
Standard Deviation	0.023
Interquartile Range	0.015
10th	0.924
25th	0.980
50th	0.990
75th	0.995
90th	0.999
Minimum	0.924
Maximum	0.999

Table 6. Plan Reliability Score Distribution for 2017 Commercial Sample

Statistic	Values
Mean	0.935
Standard Deviation	0.067
Interquartile Range	0.130
10th	0.861
25th	0.861
50th	0.953
75th	0.991
90th	0.991
Minimum	0.861
Maximum	0.991

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 considered a reasonable minimum threshold for reliability. Based on the mean reliability scores of 0.939 for Medicare, 0.982 for Medicaid, and 0.935 for commercial, the measure is considered reliable.

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, evidence- and consensus-based measure development process. This process, used in 2018-2019 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), composed 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**: **Step 3**: PQA MDTs/TFs recommend measure concepts to the PQA Quality Metrics Expert Panel (QMEP) for evaluation and refinement. The QMEP reviews and provides an initial assessment of the measure concept focusing on the criterion of importance (i.e., evidence supports that measurement can have a positive impact on healthcare quality). The QMEP votes to approve the measure concept to move forward for testing.
- **Step 4**: PQA staff prepare technical specifications (including National Drug Code [NDC] lists) for pilot testing and use 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 draft measures' feasibility and reliability and recommends whether measures should move forward for PQA endorsement consideration.
- **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.

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] 15: Initial Opioid Prescribing), 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 PQA member organizations who tested the measure using their own data, and 10 external subject matter experts.

MDT 15 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 21 members of the MDT who voted, 90.48% 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 affiliations of the 22 QMEP members in 2019 are listed in Table 7. The QMEP reviewed the measure prior to testing to ensure the importance and usefulness of the draft measure. 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 20 members of the QMEP who voted, 90% recommended that the draft measure be considered for endorsement by the PQA membership, considering the criteria of importance, scientific acceptability, feasibility, and usability.

Table 7. PQA 2019 Quality Metrics Expert Panel (QMEP)

QMEP Member Name	QMEP Member Organization
Ben Banahan	University of Mississippi

Amanda Brummel	Fairview
Lynn Deguzman	Kaiser Permanente
Marybeth Farquhar	AUA
Jessica Frank	OutcomesMTM
Shellie Keast	Mercer
Alice Lee Martin	CMS
Crystal Lennartz	McKesson
Jenny Lo Ciganic	University of Florida
Tripp Logan	MedHere Today
Jonathan Magness	Magellan Health
Jeff Pohler	Enhanced Medication Services
Dan Rehrauer	HealthPartners
Steve Riddle	Wolters Kluwer Health
Craig Schilling	AstraZeneca, LP
David Stauffer	Walgreens
Stephanie Taylor	Anthem
Christi Teigland	Inovalon
Jennifer Van Meter	Novartis
Jenny Weber	Humana
Keith Widmer	Express Scripts
Salina Wong	Blue Shield CA

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 7 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 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 82 PQA member organizations that cast a vote either in favor of or opposed to endorsement, 86.59% voted in favor of endorsing the measure.

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

The opinion of several subject matter experts was sought for input on the measure elements and assessment of the measure overall. The experts were: Kun Zhang, Health Scientist in Opioid Overdose Prevention Health Systems Team, Centers for Disease Control and Prevention; Dick Creager, Medical Director of Medical Affairs, CVS/Caremark; Kit Delgado, Assistant Professor of Emergency Medicine & Epidemiology, University of Pennsylvania; Christopher Herndon, Associate Professor in Department of Pharmacy Practice, Southern Illinois University; Jeff Schiff, Medical Director, Minnesota Medicaid; Larry Greenbalt, Professor of Medicine, Duke University School of Medicine; Don Teater, Owner, Teater Health Solutions; Aaron McKethan, Assistant Professor of Population Health Sciences and Senior Policy Fellow, Duke Margolis Center for Health Policy; Hilary Campbell, Research Associate, Duke Margolis Center for Health Policy; Nicole Brandt, Executive Director and Senior Advisor, Peter Lamy Center on Drug Therapy and Aging at the University of Maryland College of Pharmacy. All 10 subject matter experts were strongly supportive of the measure.

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.

_____

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 receiving pain management in hospice care, and those with cancer or sickle cell disease, may have unique therapeutic goals, ethical considerations, opportunities for medical supervision, and factors to consider when balancing the risks and benefits of opioid therapy. Another concern is the potential misapplication of current chronic pain management guidelines to patients outside of their intended scope. Thus, these patients are excluded from these measures whenever data are available.

The exclusions of hospice 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.

Notably, the Centers for Disease Control have also recommended unique opioid prescribing considerations for 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.

Hospice exclusions were identified using the hospice indicator from the enrollment database for Medicare and Place of Service code 34 for the Medicaid and Commercial populations. Testing was performed for the hospice

exclusion by identifying the number of patients in hospice, where available, and determining the percent of the population that would be affected by excluding patients in hospice care.

Cancer exclusions were identified using ICD-10 codes. Testing involved identifying the number of exclusions and determining the percent of the population that would be affected by excluding patients with cancer diagnoses.

Sickle cell exclusions were identified using ICD-10 codes. Testing involved identifying the number of exclusions and determining the percent of the population that would be affected by excluding patients with sickle cell diagnoses, as well as computing measure rates with and without the exclusion. Additionally, to ensure impacts of the exclusion were evenly distributed across plans, the average magnitude of contract-level change was also assessed (e.g. average difference in a given plan's score with and without the exclusion). In testing, the sickle cell exclusion was only applied to the Medicare data, as it was identified as an appropriate exclusion after testing had concluded in the Medicaid and commercial data.

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 testers 1 and 3 using the state Medicaid sample, after applying the age requirement, opioid prescription claims requirement, continuous enrollment requirement, and negative lookback period requirement, 0.35% (n=3,600) of patients were excluded based on receiving hospice and 0.33% (n=3,438) of patients were excluded based on a cancer diagnosis. As previously noted, the sickle cell disease exclusion was not included in testing for the state Medicaid sample.

For Tester 2 using the national Medicare sample, after applying the age requirements, opioid prescription claims requirement, continuous enrollment requirement, and negative lookback period requirement, 3.61% (n=1,684,154) of patients were excluded based on a cancer diagnosis, 0.012% (n=5,813) of patients were excluded based on a sickle cell disease diagnosis, and 0.62% (n=288,075) of patients were excluded based on receiving hospice care.

For tester 3 using the commercial sample, after applying the age requirements, opioid prescription claims requirement, continuous enrollment requirement, and negative lookback period requirement, <.01% (n=50) of patients were excluded based on receiving hospice and 0.47% (n=7,302) of patients were excluded based on a cancer diagnosis. As previously noted, the sickle cell disease exclusion was not included in testing for the commercial sample.

Tester 2 (national Medicare) compared measure rates with and without the sickle cell exclusion and found that changes to measure rates were negligible as noted below:

Table 8. Measure Rates With and Without Sickle Cell Exclusion

	WITH SICKLE CELL EXCLUSION:	WITHOUT SICKLE CELL EXCLUSION:
Maximum	86.6%	86.7%
Minimum	16.7%	16.7%
Mean	43.8%	43.8%

Standard Deviation	11.9%	11.9%
Interquartile Range	13.5%	13.5%

Additionally, the average magnitude of plan-level differences in performance when calculating with and without the sickle cell exclusion was found to be <0.1%.

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*)

Both testers reported low rates of all exclusions, and the sickle cell exclusion had minimal effects on measure rates. As noted previously, we believe these are important, clinically sound exclusions that improve the focus of the measure and ensure that providers are not penalized for providing appropriate care to their patients. Proper pain management is extremely important to patients, and excluded populations have unique considerations for opioid prescribing. Additionally, the exclusions are intended to improve the measure's face validity and mitigate potential unintended consequences.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

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

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

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

Not applicable.

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.

Not applicable.

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?

Not applicable.

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?
Not applicable.

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.**

Not applicable.

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*)
Not applicable.

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*)**:**
Not applicable.

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Not applicable.

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Not applicable.

2b3.9. Results of Risk Stratification Analysis:

Not applicable.

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*)
Not applicable.

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*)

Not applicable.

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 statistically significant differences in measure rates, the data described in sections above were used to calculate the mean, median, standard deviation, and interquartile range for the measure rates in the national Medicare, state Medicaid, and commercial 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. In accordance with the PQA measure specifications, measure rate calculations were only inclusive of plans with 30 or more individuals in the denominator.

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*)

The tables below show the distribution of measure rates for the Medicare population. The mean rate was 43.8%, with a median rate of 41.8%, minimum rate of 16.7%, and maximum rate of 86.6%.

Table 9. Variation in Measure Rates – 2018 National Medicare Data

Mean	Median	Standard Deviation
43.8%	41.8%	11.9%

Table 10. Distribution of Measure Rates – 2018 National Medicare Data

Statistic	Value
Minimum	16.7%
25th percentile	36.2%
50th percentile	41.8%
75th percentile	49.7%
Maximum	86.6%
Interquartile Range	13.5%
Student's t-test p-value	P<.001

The tables below show the distribution of measure rates for the Medicaid population. The mean rate was 23.7%, with a median rate of 25.9%, minimum rate of 9.5% and maximum rate of 33.5%.

Table 11. Variation in Measure Rates – 2017 State Medicaid Data

Mean	Median	Standard Deviation
23.7%	25.9%	8.1%

Table 12. Distribution of Measure Rates – 2017 State Medicaid Data

Statistic	Value
Minimum	9.5%
25th percentile	17.6%
50th percentile	25.9%
75th percentile	29.5%
Maximum	33.5%
Interquartile Range	11.9%
Student's t-test	P<0.01

The tables below show the distribution of measure rates for the commercial population. The mean rate was 25.1%, with a median rate of 24.7%, minimum rate of 23.7%, and maximum rate of 26.8%.

Table 13. Variation in Measure Rates – 2017 Commercial Data

Mean	Median	Standard Deviation
25.1%	24.7%	1.6%

Table 14. Distribution of Measure Rates – 2017 Commercial Data

Statistic	Value
Minimum	23.7%
25th percentile	24.2%
50th percentile	24.7%
75th percentile	25.7%
Maximum	26.8%
Interquartile Range	1.6%
Student's t-test	N/A*

*There were too few plans to determine significant differences between Q1 and Q3.

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 an interquartile range of 13.5%. 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 an interquartile range of 11.9%. There is a statistically significant difference in measure rates between the top and bottom quartile of the plans included in the testing (P<0.01 at alpha = 0.05). This variation shows that there are statistically significant and clinically meaningful differences in rates across plans.

For the commercial population, the measure rates did not show significant variation, with an interquartile range of 1.6%. A statistical test to determine the differences between the top and bottom quartile of the three commercial plans was not included in the testing, as the data are not appropriate for this test. However,

with only three plans in the sample, we caution against interpreting these findings as suggesting that potential for improvement does not exist in the commercial population as a whole.

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

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

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

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

Not applicable.

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

Not applicable.

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

Not applicable.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

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.

Missing data related to exclusion criteria (hospice, cancer diagnosis, sickle cell diagnosis) were not encountered, although as previously noted, only Medicare testing identified sickle cell diagnoses. However, this was not a result of missing data in the Medicaid dataset.

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.

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)

If other:

3b. Electronic Sources

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

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

ALL data elements are in defined fields in electronic claims

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

Not applicable.

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.

Measure testing indicated the measure is feasible, as the required data (prescription claims and medical claims) are readily available. No difficulties were identified.

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 of ownership to this measure and can rescind or alter the measure at any time. No use of any PQA measure is authorized without prior PQA approval of such use. All uses of PQA measures are subject to such conditions as PQA specifies, and certain uses of the measures may be subject to a licensing agreement specifying the terms of use and the licensing fee. Users of the measure shall not have the right to alter, enhance, or otherwise modify the measures.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Quality Improvement (external benchmarking to organizations)	Payment Program https://innovation.cms.gov/innovation-models/enhancedmtm Enhanced Medication Therapy Management

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

As noted in 4a1.3, CMS plans to begin reporting the IOP-LD measure in the Medicare Part D Patient Safety reports for the 2020 measurement year.
Additionally, the IOP-LD measure has been adapted for use in the CMS Enhanced Medication Therapy Management (EMTM) program. The EMTM model tests whether providing Part D sponsors with additional payment incentives and regulatory flexibilities promotes enhancements in the MTM program, leading to improved therapeutic outcomes, while reducing net Medicare expenditures. The EMTM program currently includes 2.2 million patients distributed across 22 Medicare Part D Prescription Drug Plans (PDPs). The model includes six sponsors across five Medicare Part D regions, and includes reporting by sponsor, stratified by plan. To align with the reporting frequency of EMTM monitoring measures, IOP-LD measure scores are reported by quarter as opposed to a cumulative score for the measurement year. Additionally, measures are weighed to account for the time that each beneficiary is enrolled or targeted by the Enhanced MTM plan, and the measure population is limited to EMTM enrollees.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A; See 4a1.1 and 4a1.3.

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.*)

As stated in the Advance Notice of Methodological Changes for Calendar Year (CY) 2021 for Medicare Advantage (MA) Capitation Rates and Part C and Part D Payment Policies ("2021 Advance Notice"), the Centers for Medicare & Medicaid Services (CMS) plans to begin reporting the IOP-LD measure in Part D Patient Safety reports for the 2020 measurement year, and to add the IOP-LD measure to the display page for 2023 (2021 data) and 2024 (2022 data).

CMS also notes in the 2021 Advance Notice that it "will consider adding the IOP-LD measure to the Star Ratings in the future pending rulemaking once we gain experience with the measure." Given this, PQA anticipates potential future addition of the IOP-LD measure to the CMS Star Ratings program following standard rulemaking processes.

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.

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.

As a relatively new measure, information is not available regarding performance results or feedback from measure users. However, PQA uses a consensus-based, multi-stakeholder measure development and face validity process that solicits input from a wide variety of stakeholders throughout development, including entities that will be held accountable to the measure (health plans). For more information on this process and the results, please refer to Testing Form section 1b.1 (Validity Testing).

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.

PQA makes measure specifications available to its members before PQA endorsement votes in addition to holding a 3-week comment period and hosting an educational webinar on the measure.

Please also see Testing Form section 1b.1 for further detail on multistakeholder engagement and input during measure development.

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.

During the PQA endorsement process, comments received from a variety of stakeholders, including health plans, indicated support for the measure while raising key questions about the potential for a sickle cell disease exclusion, which was not originally included but is now included in the measure as specified in this submission.

Please also see Testing Form section 1b.1 for further detail on multistakeholder engagement and input during measure development.

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

During the PQA endorsement process, comments received from a variety of stakeholders, including health plans, indicated support for the measure while raising key questions about the potential for a sickle cell disease exclusion, which was not originally included but is now included in the measure as specified in this submission.

Please also see Testing Form section 1b.1 for further detail on multistakeholder engagement and input during measure development.

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

During the PQA endorsement process, comments received from a variety of stakeholders, including health plans, indicated support for the measure while raising key questions about the potential for a sickle cell disease exclusion, which was not originally included but is now included in the measure as specified in this submission.

Please also see Testing Form section 1b.1 for further detail on multistakeholder engagement and input during measure development.

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.

In response to stakeholder comments regarding a sickle cell disease exclusion, PQA presented the potential exclusion to the PQA Measure Update Panel (MUP) for consideration. The MUP is a standing panel of 20-25 subject matter experts responsible for making maintenance recommendations on existing PQA measures. The MUP voted in favor of the exclusion, which was then tested and presented to the PQA Quality Metrics Expert Panel, who voted and provided final consensus-based approval for the addition of the sickle cell disease exclusion.

Please also see Testing Form section 1b.1 for further detail on multistakeholder engagement and input during measure development.

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.

This is a new measure that has only recently been implemented, and therefore data on improvement over time are not available. However, we anticipate that the performance results can be used to establish benchmarks and identify opportunities to reduce initial opioid prescribing for long duration. As noted in the evidence form and measure rationale, a significant body of evidence finds that the duration of initial opioid exposure is associated with a higher likelihood of long-term opioid use, which is linked to greater risks of abuse, overdose, and other negative outcomes. Furthermore, as demonstrated in section 1b.2, there is a significant opportunity for improvement, with mean measure rates among various lines of business ranging from 23.7% to 43.8%.

Evidence has demonstrated that a variety of educational and quality improvement programs can encourage safer opioid prescribing, although literature has generally not focused on initial opioid prescribing specifically. Liebschutz et al. found that a multicomponent intervention, including a nurse care manager, electronic registry, data-driven academic detailing, and electronic decision tools resulted in improved adherence to guideline-concordant care for patients with chronic pain on long-term opioid therapy in a cluster-randomized clinical trial among 53 primary care physicians. Patients of intervention physicians were also more likely to see a 10% dose reduction (OR, 1.6, 95% CI 1.3-2.1, p<.001).(1) Wong et al. found that use of electronic pain and opioid management templates, workflow redesigns, and RN pre-visit planning and physician-nurse huddles before visits within an internal medicine residency clinic resulted in decreases in average daily MME from 96.6 to 67.7 (P<0.0001).(2) Seal et al. found that use of an Integrated Pain Team within VA primary care settings (intervention N=147, control N=147) resulted in greater morphine equivalent daily dose reductions (42mg vs. 8 mg after 3 months, 56 mg vs. 17 mg after 6 months, p<.01), as well as significant improvements in opioid risk mitigation by 6 months, including decreased co-prescription of opioids and benzodiazepines.(3)

1) Liebschutz JM, Xuan Z, Shanahan CW, et al. Improving Adherence to Long-term Opioid Therapy Guidelines to Reduce Opioid Misuse in Primary Care: A Cluster-Randomized Clinical Trial. JAMA Intern Med. 2017;177(9):1265-1272. doi:10.1001/jamainternmed.2017.2468

2) Rachel Wong, MD, MPH, William Carroll, MD, Astha Muttreja, MD, Victor Garcia, MD, Erin Taub, MPH, Alice Fernan, RN, Improving Opioid Management and Resource Utilization in an Internal Medicine Residency Clinic: A Before-After Study over Two Plan-Do-Study-Act Cycles, Pain Medicine, Volume 20, Issue 10, October 2019, Pages 1919-1924, https://doi.org/10.1093/pm/pny239

3) Seal, K.H., Rife, T., Li, Y. et al. Opioid Reduction and Risk Mitigation in VA Primary Care: Outcomes from the Integrated Pain Team Initiative. J GEN INTERN MED (2019). https://doi.org/10.1007/s11606-019-05572-9

4b2. Unintended Consequences

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

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

This is a new measure that has not yet been implemented, and therefore there are no unexpected findings from implementation to report. PQA notes that the IOP-LD measure is intended for use for retrospective

population-level performance measurement, and is not intended to guide clinical decision-making for individual patients.

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

This is a new measure that has not yet been implemented, and therefore there are no unexpected findings from implementation to report.

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

3389 : Concurrent Use of Opioids and Benzodiazepines (COB)

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

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

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.

Most of the PQA opioid measures (NQF # 2940, 2950, 2951, and 3389) use the same target population (denominator), and each have different areas of focus (numerator) related to opioid prescribing. The PQA AMO measure (NQF #3541, recommended for endorsement by the Behavioral Health and Substance Use Standing Committee and awaiting CSAC approval) shares a related denominator, but includes only individuals on long-term opioid therapy and has a different area of focus related to drug testing. The NCQA opioid measures were developed as an adaptation to existing PQA measures; the NCQA opioid measure denominators are similar to the PQA opioid measures but have a different area of focus than the IOP-LD measure.

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 address 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): Pharmacy Quality Alliance

Co.2 Point of Contact: Lynn, Pezzullo, lpezzullo@pqaalliance.org, 401-474-9706-

Co.3 Measure Developer if different from Measure Steward: Pharmacy Quality Alliance

Co.4 Point of Contact: Lynn, Pezzullo, lpezzullo@pqaalliance.org, 401-474-9706-

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.

The roster for the Measure Development Team (MDT) involved in development of the IOP-LD measure is below. PQA MDTs, composed 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 focus on specific aspects of the medication-use system and/or specific therapeutic areas and benefit by having their development work reviewed by larger stakeholder groups. They may also receive input from other PQA panels, and recommend measure concepts to the PQA Quality Metrics Expert Panel (QMEP) for evaluation and refinement. For more information, please refer to the Testing Form, section 2b1.2.

John Beckner / National Community Pharmacists Association

Elizabeth Bentley / Kaiser Permanente

Jonathan Bosold / SinfoniaRx

Hilary Campbell / Duke University

Dick Creager / CVS Caremark

Victoria Dang / Centers for Medicare & Medicaid Services

Jose Diaz / Express Scripts

Rachel Digmann / Telligen

Joel Farley / University of Minnesota

Meron Gartner / Outcomes MTM

Travis Gau / Genoa Healthcare

Brad Gregory / Optum

Genevieve Hayes / Medical University of South Carolina

Zachariah Hicks / Rite Aid

Diana Higgins / Veteran's Affairs

Hank Hoang / Health Resources and Services Administration (HRSA)

Lindsay Joseph / University of Pennsylvania Medical Center

Richard Logan / Semo Rx

Michael Long / Federal Board of Prisons

Shelly Nance / Kroger

Lauren Narkiewicz / Wellcare

Patricia Neafsey / ActualMeds

Alpa Patel / Aetna

Sujith Ramachandran / University of Mississippi

Devanshi Sheri / Amerigroup

Judy Sommers Hanson / Walgreens

Kathleen Vest / Midwestern University

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2019

Ad.3 Month and Year of most recent revision: 03, 2020

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

Ad.5 When is the next scheduled review/update for this measure? 03, 2021

Ad.6 Copyright statement: Rights retained by PQA Inc, 2020.

Ad.7 Disclaimers: The IOP-LD measure is intended for use for retrospective population-level performance measurement, and is not intended to guide clinical decision-making for individual patients.

Ad.8 Additional Information/Comments: