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 #: 3615

Corresponding Measures:

De.2. Measure Title: Unsafe Opioid Prescriptions at the Prescriber Group Level

Co.1.1. Measure Steward: Centers for Medicare and Medicaid Services

De.3. Brief Description of Measure: Percentage of all dialysis patients attributable to an opioid prescriber's group practice who had an opioid prescription written during the year that met one or more of the following criteria: duration >90 days, Morphine Milligram Equivalents (MME) >50, or overlapping prescription with a benzodiazepine.

Please note that the opioid prescriber is the clinician identified from Part D Medicare Claims who actually provides an opioid prescription to a dialysis patient. This provider is usually not the nephrologist who is overseeing the patient's dialysis care. This is in contrast to NQF submitted measure #3616, which is at the dialysis provider level (the clinician who receives the Monthly Capitated Payment for overseeing dialysis care). While the dialysis provider is usually not the clinician who is prescribing opioids, the MCP physician does have a responsibility to be aware of dialysis patients' medications and that doses are safe and appropriate for level of kidney function.

The proposed measure is a directly standardized percentage, which is adjusted to the national distribution of covariates (e.g., age, gender, risk factors). Here, "national" refers to all opioid prescriber groups combined. Specifically, the standardized rate for a given prescriber's group is an estimate of the group's percentage of unsafe opioid prescriptions if their case-mix were equal to that of the national population. Case-mix adjustment is based on a logistic regression model.

1b.1. Developer Rationale: Several observational studies have demonstrated an association between unsafe opioid use in the dialysis population and higher risk of fall/fracture, hospitalization, and mortality. Unsafe opioid use is typically defined as >50 morphine milligram equivalents (MME), duration > 90 days, or co-prescription with a benzodiazepine.

The measure focus is the process determining the percentage of all dialysis patients attributable to an opioid prescriber's group practice who had an unsafe opioid prescription written within the year.

The measure is risk adjusted to mitigate against the unintended consequences of under treatment of pain in patients with comorbidities that have a significant pain component (e.g., cancer, sickle cell disease). By adjusting for case-mix at the prescriber's group practice, our intent is for providers to be able to write necessary opioid prescriptions for patients with greater comorbidity, and likely greater analgesia needs, since

the measure does not penalize individual prescribing events or hold providers to an absolute scale or threshold. Rather, the measure identifies the small number of group practices that, based on their year-long prescribing patterns, have extreme deviations relative to their peers.

Once implemented practitioner performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in reducing unsafe opioid use, while accounting for patients where higher dose or longer-term therapy may be warranted.

S.4. Numerator Statement: The numerator is the number of patients in the denominator who were prescribed an opioid that was either >90 days duration during the year, >50 MME, or overlapped in time with a benzodiazepine prescription.

S.6. Denominator Statement: The denominator is the number of patients associated with an opioid prescriber's group practice who are receiving maintenance dialysis (in-center or home dialysis) for any duration who receive an opioid prescription during the one-year reporting period.

S.8. Denominator Exclusions: Patients who have a hospice claim at any time (either before or after the opioid prescription date) during the one-year reporting period are excluded.

De.1. Measure Type: Process

S.17. Data Source: Claims, Other, Registry Data

S.20. Level of Analysis: Clinician: Group/Practice

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

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

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

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

Preliminary Analysis: New Measure

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

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 included "Other Source of Evidence" to support the development of this process measure. The development of this measure wasn't based on clinical practice guideline, US Preventive Services Task Force Recommendation (USPSTF), or systematic review and grading of the body of evidence.
- The developer provided a logic model to show that several observational studies have demonstrated an association between unsafe opioid use in the dialysis population and higher risk of fall/fracture, hospitalization, and mortality.
- The developer also provided empirical evidence from the literature to link unsafe opioid prescription practices to serious adverse event, such as hospitalization and mortality, in the dialysis population. Furthermore, interventions such as use of PDMPs and co-prescription of naloxone have been demonstrated to reduce these risks.
    - The developer provided the search terms/query that was conducted in PubMed in February 2019, which yielded 268 articles that were reviewed and of these 43 were selected for presentation to the Technical Expert Panel that was convened to make recommendations regarding this measure. The developer provided a list of references for relevant articles and a summary synthesizing the evidence in the testing form.

Exception to evidence

- NA

Questions for the Committee:

- *For structure, process, and intermediate outcome measures:*
    - *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?*
    - *Does the Standing Committee agree that the submitted evidence indicates **high certainty** that benefits clearly outweigh undesirable effects?*
- For possible exception to the evidence criterion:
    - Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?
    - Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?
    - Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empirical evidence?

Guidance from the Evidence Algorithm

Process measure without systematic review (SR) or grading of the body of empirical evidence (box 3) à empirical evidence submitted but without SR and grading of the evidence (box 7) à Summarized empirical evidence includes all studies in the body of evidence (box 8) à **Moderate**

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

RATIONALE:

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 provided January 2017 – December 2017 data analysis, which indicates the physician-level mean percentage of patient months with unsafe opioid use is 39.7%, which standard deviation of 19.8%, median of 38.5 and interquartile range extending from 25% up to 52.6% (N of prescriber groups=5,123, N of patients= 204,034).
    - The developer noted that of the ESRD patients who are prescribed an opioid, 39.7% of those prescriptions met the above definition for unsafe use.
- The data provided demonstrates that there is substantial variation in provider group performance indicating that a performance gap exists that may be modifiable.

Disparities

- The developer evaluation data from January – December 2017 in a logistic regression model for unsafe opioid use and provided odds ratios for the patient characteristics including age, sex, race, ethnicity, dialysis vintage, employment status, Medicare coverage, and Area Deprivation Index (ADI).
    - Data on patient level SDS/SES factors were obtained from Medicare claims and administrative data; zip code level data for the Area Deprivation Index (ADI) are obtained from Census data (2009-2013), based on patient zip-code.
- Based on the data, the developer concluded that age, sex, race, and ethnicity are all statistically significant predictors of unsafe opioid use.
    - Patients under age 25 had 17% higher odds of having unsafe opioid use for each year increase in age.
    - Between ages 25 and 65 the odds decreased with age (0.7% each year) with a sharper decline after age 65 (2.7% each year).
    - Females had 5% higher odds of having unsafe opioid use versus males.
    - Hispanic ethnicity was associated with lower odds of Opioid unsafe use whereas Black race had 35% lower odds of unsafe opioid use compared to whites.
    - Unemployment or "other" employment status as well as dual eligible status were all associated with higher odds of Opioid unsafe use.
- The developers noted that the analysis results for age, race, sex and patient SES indicate potential disparities in unsafe opioid use.
- Additionally, the developers stated that patient-level SDS/SES variables are not included as adjustments in the measure since, in the absence of biological effects explaining these differences, risk adjustment for these factors could potentially mask disparities in care. However, these variables do highlight certain subgroups that may be at higher risk for unsafe opioid use as prescribers consider interventions to close performance gaps.

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

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

Measure scores by tertile showed a trend towards higher hospitalization rate and hospital days (not risk adjusted). Mortality rate differences were slight but diverged more strikingly when looking at 30 d intervals over a 12 month period. No statistical significance is assigned and it is not clear whether patients were receiving opioids in the assigned month. Pertinent outcomes such as treatment for overdose, falls, fractures are noted.

evidence was based on observational studies and literature review. USRDS data sited was from 2010 and CDC guidelines were from 2016. This is a process measure holding dialysis physicians responsible for the prescriptions of other providers. The evidence relating to hospitalizations and mortality being linked to opioid use versus underlying causes is not presented.

Process Measure. Measure outcomes to decrease dialysis patient falls, fractures, hospitalizations and decrease mortalities. Does apply to desired outcomes. Applied directly. Logic model used to show that several observational studies have demonstrated an association between the Prescription Drug Monitoring Program and co-prescription of naloxone have been demonstrated to decrease those risks.

Evidence supports measure

Process Measure. Measure outcomes to decrease dialysis patient falls, fractures, hospitalizations and decrease mortalities. Does apply to desired outcomes. Applied directly. Logic model used to show that several observational studies have demonstrated an association between the Prescription Drug Monitoring Program and co-prescription of naloxone have been demonstrated to decrease those risks.

No systematic review or quality, quantity and consistency of evidence. Development was not based on clinical practice guidelines or US Preventive services Task Force. Developer provided a logic model and evidence from the literature to link unsafe opioid prescription practices to adverse events including mortality in the dialysis population

Satisfactory evidence and rationale

The evidence provided seems to be directly related to the process being measured. However, one of the frequently quoted sources states that a causal relationship cannot be inferred. Additionally, it noted that opioid prescriptions may be indicative of more severe illness. Moreover, pain has been found to be frequent in people with ESRD and may impact QOL.

Indirect evidence to support the measure

There is moderate evidence that the measure would directly apply and the negative outcomes of falls, morbidity, and mortality would decrease.

yes

Very hard to link the measure as constructed to the quality of care provided by the dialysis facility or dialysis provider (see comments below on reliability and validity)

moderate supporting data at best

There is extensive evidence demonstrating a correlation between opioid use and poor outcomes in CKD-5 patients

The evidence is directly applicable but it is three years old. There are only correlational data not causation data. Prior to obtaining data from the measure we cannot say if patients receive opiates because they have multiple co-morbid conditions which predispose to hospitalization and death or if the opiates cause increased rates of hospitalization and death. New information can be obtained from newer Medicare data sets for better understanding of prescribing patterns and their relationship to hospitalization and mortality.

Low - no systematic review and all observational data

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

Yes (if 2017 is considered current). Interquartile range of patient months with high risk rx extends from 25% to 52.6. This reflects variability in care. Age, sex, race, SES results suggest disparities exist.

Would like more current data on opioid prescribing practices to be able to determine gaps in care. Data was provided by populations subgroups.

No current performance gap data measured,

Gap is documented.

No current performance gap data measured,

Age, sex, race and ethnicity all predictors of unsafe opioid use as is unemployment.

2017 data used that shows substantial IQ range that supports a performance gap; some disparity data provided

Yes, performance data were provided and a gap was shown. Data revealed that the physician level mean score was 46.5% and that there was performance variation across providers ranging from scores of 0 to 92.3%. Data by population subgroups suggested older age, sex, race, and ethnicity are all statistically significant of unsafe opioid use.

Yes, data were provided and gap in care of unsafe opioid practice as well as disparities in subgroups are noted.

There is a moderate performance gap but may be modifiable with prescriber motivation to consider other interventions to close the gap.

Yes; appropriate

(see comments below) the comment that approximately 39% of HD patients are receiving "opioids in unsafe doses" may be linked to prescribing patterns and secondarily to the data definitions within the measure. the performance gap noted may reflect that same issue.

Yes, data provided, appears there is a performance gap, but not clear if they represent care disparities

The performance gap does warrant a national performance measure. The evidence demonstrates a unsafe use of opioids in CKD - 5 patients..

No. Only 3.5% of provider groups have unsafe opioid prescribing practices. It is not known if these data are consistent over time or represent one-time variation. It is not clear if this demonstrates disparities in care.

Moderate - gaps exists based on presented data

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

2a. Reliability

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

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

2b.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**

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by the NQF Scientific Methods Panel (SMP) and discussed on the call. The Subgroup passed the measure on reliability and validity. The measure was pulled for discussion during the March 2021 SMP meeting. A summary of the measure and the Panel review and discussion is provided below.

Reliability

- The SMP passed the measure on reliability with High rating (H-6; M-1; L-1; I-1).
- The developer conducted validity testing at the performance measure score level using inter-unit reliability (IUR) for the annual performance scores.
- The developer used CROWNWeb, Medicare Claims, the CMS Medical Evidence form 2728, Medicare Part D Claims as data sources to test the measure. The analysis included 103,157 physicians in 5,123 groups (range: 1-2,328 clinicians) with an average of 40 patients per group (range: 11-2,411).
- Physician groups must have more than 10 eligible patients to be included in the measure or the analysis.
- The developer noted that the IUR calculated at the group level is 0.86 which means 86% of the total variation of this prescriber group level measure can be explained by the differences among prescribers and not by random noise.
- To assess further whether the measure can identify prescriber groups with extreme values, we computed the Profile inter-unit reliability (PIUR), which is 0.98. The developer stated that the discrepancy between the IUR (0.86) and PIUR (0.98) indicates the existence of outlier prescriber groups that can be identified by the measure.

Validity

The SMP passed the measure on validity with Moderate rating (H-2; M-4; L-1; I-2).

Validity testing was conducted at the score leve*l:*

1. The developer conducted a concordance analysis of the relationship between measure scores, hospitalization, and mortality.
2. Hospitalization rate at the practitioner group level is 1.49, 1.46 and 1.41 for T1, T2, and T3 respectively (trend test $p<0.001$), while the average number of hospital days per year and patient at the practitioner group level is 6.1, 5.1 and 4.1 respectively (trend test $p<0.001$).

3. The practitioner group level average mortality rate is 0.19, 0.20, and 0.18 per patient-year for T1, T2, and T3 groups, respectively.

- SMP Subgroup pilled this measure for discussion specifically to address an overarching question: To what extent is the validity analysis confounded by unmeasured case mix, considering that dialysis physicians with sicker patients (e.g., those with comorbid cancer) have higher mortality rates, hospitalization rates, and opioid use. The two measures were therefore discussed concurrently.
- During the SMP meeting, concerns were raised regarding the use of a risk adjustment model for a process measure. They noted it would be more appropriate for risks to be made into exclusions (e.g., cancer); and the other factors that are endogenous (e.g., drug dependence, substance use disorder, anxiety disorders, and previous opioid poisoning) may increase risk and are confounders that may be difficult to understand or differentiate. The risk adjustment model was noted as appropriate in terms of performance statistics but lacked an underlying theory to justify the selection of factors for the model.
- The SMP also expressed concerns that the validation of the measure is based on dividing provider groups into tertiles that showed the top tertile with a failure rate over 46 percent, the middle at 30-36 percent, and the best tertile under 30 percent. The submission noted that patients in the worst performing tertile have a slightly higher hospitalization odds ratio, 1.49 versus 1.41 and a few more hospital days per year, 6.1 versus 4.1, as well has a higher death rate. They also noted these findings were reported under an unadjusted analysis when the developer has suggested that risk adjustment is essential for the measure's application.
- The SMP elected not to revote on the measure but passed along the concerns to the Renal Standing Committee. A full summary of the SMP discussion is linked here.

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 Scientific Methods Panel 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 Scientific Methods Panel expressed concerns with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

Committee Pre-evaluation Comments:

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

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

- It is not clear how the patient gets attributed to a particular provider if more than one prescribes opioids during the year. Does the 90 day criterion require consecutive days? It is not clear that the case mix adjustment is always. biologically plausible.
- The MME.50mg is not well defined. Is it for the 90 day period. Hospice needs clarification.
- Moderate rating…used inter-unit reliability
- Reliable specifications
- Moderate rating…used inter-unit reliability

- Data elements are repeatable and score is precise to distinguish differences in performance across providers
- unclear as to how multiple prescribers of opioids to a single pt are assessed -- for instance, if >90 days of opioid provision is a result of several prescribers, how would the first prescriber (who may have provided a short course of drug) know that someone or multiple prescribers in the next few weeks would later provide additional opioid that add up to a longer course and "dangerous" designation?
- Data elements were provided and clearly defined. Specs and the calculation algorithm are clear. The measure should be able to be consistently implemented.
- No concerns. Data elements seem to be clearly defined.
- I have no concerns.
- none
- several concerns regarding reliability and validity. With respect reliability unclear how hospice enrollment will be defined. It is stated that this will come from the CMS file? greater specificity regarding the specification of that file should be included. When /how in the course of dialysis patients' plans of care is that information added to their CMS database? 2. The assignment of patients based on the prescribing pattern of primary care providers, and pain management providers within one group practice , a priori, to particular dialysis provider or dialysis facility is not necessarily valid. As developed this construct presumes that the primary care provider/pain provider and dialysis provider are all linked together in the care of the patient because they are all linked together in one provider group. This is not necessarily the case, and must be validated for this measure to be reliable. see comments below
- no concerns
- Not clear when multiple practitioners prescribe opioids - who would then be identified as the "unsafe" prescriber?
- Data elements are clearly defined.
- No issues

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

- No
- What constitutes "unsafe prescribing practices may be variations depending on diseases.
- No concerns
- IUR/PIUR indicates reliability
- No concerns
- no
- IUR and PIUR both very high
- The IUR was 0.60 which is acceptable reliability.
- No concern about reliability
- No.
- no
- see below
- no concerns
- no - IUR is 0.86
- No. Tests for reliability show adequate response.
- No issues
- 2b1. Validity -Testing: Do you have any concerns with the testing results?

- Risk adjusted data were not used apparently in assessing validity. Mortality results are difficult to interpret. The description of an unsafe prescription is not clearly aligned with the literature. Risk of opioids is continuous. The comparison is really between lower and higher risk dosing.
- As stated above
- no concerns with testing results since it is data collection Scientific Methods Panel is satisfied with the validity
- no
- no concerns with testing results since it is data collection Scientific Methods Panel is satisfied with the validity
- no
- performance on measure compared to hospitalization rates and duration and mortality data with poor performance on this measure found to track with higher hospitalization rates for longer periods of time, but no way to assess if confounding variables in play that would associate all of these and no correlations provided
- Yes, there are concerns. Needed information is missing. Validity was tested at the performance measure scores by evaluating concordance between measure scores, hospitalization metrics and mortality rates. Looking at mortality, to account for potential selection bias stemming from the definition of chronic opioid use requires the patient to live at least 90 days, patients were instead stratified based on length of time at risk during the 12- month performance period. It is not indicated how the time at risk stratification was performed. P-values are also not included for mortality stratification so one doesn't know if the results are statistically significant.
- may need to address other comorbid conditions which may affect validity
- No
- none
- Many large primary care groups care for patients over a very wide geographic area. and some practice (care) groups may have hundreds of providers, (all within one practice TIN number). This measure assumes that every ESRD patient given opioids by the primary care/pain providers of that group will be under the care of a nephrologist who is also a member of that group, and that the patient receives dialysis care under the auspices of that nephrologist. As developed, the measures automatically presumes that the primary care provider, and the nephrologists are linked such that if the primary care provider is using opioids incorrectly the dialysis provider (nephrologist) should notice this when they see the patient and correct it. Failure to do so will result in a negative outcome for this measure. However, the NEPHROLOGIST in the practice group MAY NOT be at all involved in the care of that patient: the patient may be under the care of an entirely separate nephrologist and be receiving dialysis in a completely separate dialysis facility. In fact, in some geographic areas that would be a very common scenario, wherein It is just as likely that patients seen by the primary care or pain members of a practice group will be receiving specialty care forma completely different group and seeing a nephrologist who is not a member of the practice group. As the measure is constructed the nephrologist who IS a member of the practice group will be cited for inappropriate care. The frequency of this scenario needs to be assessed in order for the validity of this measure to be ascertained. 3. The EXACT definition of " greater than 90 days of opioid use " needs to be carefully clarified. Is that based on prescriptions FILLED or prescriptions WRITTEN (as tracked in PDMPs available in some, but not all, states). As defined in the measure, low-dose opioid use (any use at all) for over 90 days within a year is classified as "unsafe opioid use," in this measure. Thus, a patient could receive a SINGLE prescription for four times a day use for 30 days, but only take the medication once a day and have enough pills to last them for more than 90 days. Contrast that with another patient situation wherein the practitioner may write 4 separate scripts for small doses for 30 days each. Which one of these patterns will be classified as a "fall out" by the measure as "unsafe opioid

use?" either / neither or both.? Thus, the specificity of exactly how opioid prescribing and utilization will be tracked is critical for this measure.

- Concerns that prescriber group level makes the dialysis prescriber responsible for the actions of other care providers. Even if the dialysis prescriber knows of the prescription, the ability to influence the other prescribes is likely overstated.
- The hospitalization rates and mortality rates could be affected by other factors. But good start.
- Risk adjustment is inadequate; it does not account for co-morbid medical conditions.
- No issues

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

- Only gender was used for SDS/SES adjustment. Biological rationale for risk adjusted not always is clear
- Exclusions are not complete. Need to adjust for co-morbid conditions that could impact opioid use. No definition regarding time on dialysis. Data through 2010 but no data on trends after 2010.
- No
- Revision of risk adjustment may be indicated per SMP
- No
- Exclusions are consistent with the evidence. Physician groups must have more than 10 eligible patients to be included in the measure. There may be a risk of double counting patients. IUR was 0.86
- no obvious threats
- Exclusions: the exclusions may not be sufficient to prevent under treating patients in pain where long-term use of opioids is warranted. Only hospice patients are specifically mentioned. Patients suffering with debilitating chronic pain and those with potentially terminal conditions who are not eligible for hospice yet should be excluded as should patients in palliative care and under the treatment of a pain management specialist. Risk adjustment: It is unusual for process measures to be risk-adjusted. CMS states that process measures are not risk-adjusted; rather the target population of a process measure is defined to include all patients for whom the process measure is appropriate. If risk adjustment is necessary, should this measure, as well as the other measure, be classified as intermediate outcome measures?
- appropriate to exclude hospice patients; may need to consider other exclusions as comorbid conditions could skew data
- n/a
- yes
- Parts of the measure rationale note that the risk adjustments will be utilized. However, enrollment in a hospice is the only exclusion criteria and is unclear how the risk adjustments delineated will be applied in a process measure. Additionally, much of the data regarding the risk of mortality and hospitalization did not use a risk adjusted model and I am unclear how risk adjustments will be applied. the fact that 39% of patients on dialysis were deemed to meet the criteria for "unsafe opioid use" may be related to the definitions used in the measure. As noted above, ANY opioid use for over 90 days aggregated in

any 365-day period is viewed as unsafe use. Appropriate risk modeling, and adjustment for co-morbid conditions and clinical events, may influence that classification.

- I again feel that the exclusion criteria limited to hospice is quite limiting.
- there are extensive variables, with social variables eliminated. But may not affect overall results.
- There are differences in opioid prescribing patterns based on SDS factors which are not included in the model.
- Limited SDOH data in the data sets used makes determining the impact of such factors on this measure difficult. Data also relies on Medicare FFS patients and so MA and private pay is missing.

2b4-6. Other Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- Since it is unclear what is driving outlier prescription activity - poor quality lack of risk adjustment? Missing data have minimal impact
- Not sure the measure identifies meaningful differences regarding quality as the measure is defined. Need more clarification regarding the data about overlapping with the benzodiazapams
- Missing data does constitute a threat to the validity of this measure.
- No
- Missing data does constitute a threat to the validity of this measure.
- concerns were raised regarding the use of a risk adjustment model for a process measure. They noted it would be more appropriate for risks to be made into exclusions (e.g., cancer); and the other factors that are endogenous (e.g., drug dependence, substance use disorder, anxiety disorders, and previous opioid poisoning) may increase risk and are confounders that may be difficult to understand or differentiate.
- no other significant threats identified
- 2b4 Meaningful differences: For each provider group the proportion of patient-months with a high-risk prescription was calculated at the year-level and then compared to the overall national distribution. 90-93% of providers' were categorized "as expected". The developer interpreted this as demonstrating "both practical and statistically significant differences in performance". This interpretation is questionable. Does a measure where 90% of providers are already performing "as expected" justify addition to publicly reported accountability programs? Would it provide enough meaningful, actionable information to stakeholders to support its use? 2b5 Comparability of performance scores: NA. 2b6 missing data/no response: NA.
- missing data could be a threat to validity
- I see no threats to validity. The three tertiles the data was closely comparable although not adjusted for case mix.
- no
- none noted
- only Medicare with D
- Did not address overall missing data.
- It is unclear if the differences are meaningful. The proportion of provider groups performing outside the expected performance is small and misclassification is a risk.
- Non Medicare FFS data will be missing

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

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

All data elements are in defined fields in a combination of electronic sources.

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?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- Not clear how Part D med claims will match up with individual provider understanding of whether patient falls into unsafe category. If there is more than 1 prescriber and inconsistent prescribing over time, it may be hard to recognize high risk prescribing.
- No concerns
- No concerns. feasibility ryes data already inputted in registry.
- No concerns
- No concerns. feasibility ryes data already inputted in registry
- Data collected by healthcare personnel during the provision of care in defined fields in electronic records
- all data elements required in EHR or other electronic sources in defined fields
- No feasibility concerns.
- no concerns
- No concerns
- None
- as above the exact data and definition of opioid prescriptions (number of pills per scripts, numbers of refills) need clarification. The assignment of patients to a nephrologist based on the activities of the primary care provider in that same group needs validation
- no concerns
- The data is easily accessible. Care must be provided on what category the data should apply.
- Patients are required to participate in Medicare part D. This may introduce bias since patients with more expensive prescriptions may be more likely or less likely to opt into Medicare part D

- No issues

Criterion 4: Usability and Use

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

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

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

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

Current uses of the measure

Publicly reported? ☐ Yes ☒ No

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

OR

Planned use in an accountability program? ☐ Yes ☐ No

Accountability program details

The developer stated that NQF #3515 is a newly submitted measure that is currently undergoing initial endorsement review.

Additionally, the developer noted that CMS will determine if/when to report this measure in a public reporting/payment program. One potential application for the measure is in the Quality Payment Program where it would be one of several optional measures that a group practice could select in their self-evaluation.

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

N/A (new measure)

Additional Feedback:

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

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

Improvement results

The developer noted that since the measure has not yet been implemented in the public program, improvement data isn't available. Additionally, developers stated that CMS currently anticipates implementation of this unsafe opioid measure. Once implemented prescriber performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in reducing unsafe opioid use, while accounting for patients where higher dose or longer-term therapy may be warranted.

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

Unexpected findings (positive or negative) during implementation

N/A (new measure)

Potential harms

N/A (new measure)

Additional Feedback:

N/A (new measure)

***Questions for the Committee**:*

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

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

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

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

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

- N/A
- Not currently publicly reported
- Not being publicly now. CMS will determine when to report in public reporting/payment application for the measures in the Quality Payment Program. public reporting/payment application
- Not currently publicly reported
- Not being publicly now. CMS will determine when to report in public reporting/payment application for the measures in the Quality Payment Program. public reporting/payment application

- New measure not currently publicly reported or part of any accountability program. CMS will determine if/when to report to the public
- new measure not used thus far
- 11.4a1 Accountability and Transparency: Yes. A credible plan for implementation was provided. 11.4a2 Feedback on measure: A public review and comment period was provided by the developer earlier this year.
- no comments
- The data in this measure is not currently calculated but is on CMS radar for an accountability program
- no concerns
- see below
- as far as I know, no feedbacks currently being given on this measure
- No issues.
- CMS will have to decide how to report the data. Unknown.

Developer conducted a TEP, other than that, no feedback from HCPs is noted in the application.

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

- Since opioid risk may be continuous, if measure results in reduction in prescribing that might be good. Measure may set up provider patient conflict, solicitation of meds from other prescribers.
- Benefits of this measure do not outweigh the potential harm to patients with chronic/persistent pain.
- benefits of measure include high quality and efficient healthcare outweigh any unintended consequences. Measure used to decrease falls, fractures, hospitalizations and decrease mortality,
- Highly usable measure
- benefits of measure include high quality and efficient healthcare outweigh any unintended consequences. Measure used to decrease falls, fractures, hospitalizations and decrease mortality,
- New measure and unable to assess
- seems usable without unintended consequences
- 12.4b1 Usability - improvement: Yes. A logic model was included in the materials. 12.4b2 Usability - Benefits vs. harms: I believe that the risks of unintended consequences are considerable. In its February 2020 report, a NQF TEP that reviewed quality measures related to opioids recommended that opioid measures used in Federal quality programs should address any of a number of patient-centric clinical issues. Examples included recovery from opioid use disorder, assessment and treatment of physical and mental health comorbidities, co-prescription of naloxone, patient-centered analgesia, and appropriate opioid tapering. The two measures submitted for our review do not address any of these recommendations. They focus instead on reducing opioid use and appear to disregard clinical decision-making, as well as the etiology or severity of the pain being treated and quality of life issues. To further improve pain management, the epidemiology of pain in patients on dialysis, as well as patients' unique needs and prefer
- under treatment of pain
- The performance measure goal would be to use the data to motivate prescribers to use other methods for pain management in CKD patients. Yes, the benefits outweigh the harm.
- none

- inappropriate opioid use is a critical concern among patients with CKD and ESRD. The indiscriminate use of opioids within this population warrants close attention and alternative measures of pain control must be evaluated and best practices created. The Kidney Disease Hemodialysis Opioid Prescription effort (HOPE) consortium should provide insight into best practices for opioid management in this complex population. Based on the data suggesting that upwards of 39% of dialysis patients are using opioids in an " unsafe manner," it is important that alternative safe, and effective therapies be readily available for these patients to avoid other unintended consequences as may occur with their sudden discontinuation.
- I think this measure may punish individual providers based on the actions of others in the group.
- This measure could be extremely beneficial. Ensuring the data is correctly analyzed is critical to make this measure viable.
- Unclear. The developers state they do not know if the measure is useful for changing prescribing patterns.
- Risk is inadequacy pain control for ESRD patients.

Criterion 5: Related and Competing Measures

Related or competing measures

The developers did not identify any related or completing measures for NQF #3515.

Harmonization

N/A

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

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

- No
- N/A
- No related or competing measures
- N/A
- No related or competing measures
- No; 3616 will be related
- N/A
- NA
- no
- No
- none
- no current competing measures
- related measure 3616
- none
- 3616
- None that are endorsed.

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 6/8/2021

Comment by: The American Medical Association (AMA)

The American Medical Association (AMA) appreciates the opportunity to comment on this measure. We have significant concerns as we believe that it is not aligned with the evidence as specified and there are significant unintended negative consequences that could be experienced with its use. 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.

The ongoing singular focus on the dose and duration of opioid prescriptions disregards the important steps that have already been 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 and the CDC stated that its guideline should not be used to create hard and fast policy. In fact, the CDC is currently in the process of updating the guideline and the AMA provided in-depth feedback on our concerns to the CDC during last year's public comment (AMA, 2020).

The AMA disagrees with the fundamental premise of measures that focus on daily dose and duration of therapy involving prescription opioid analgesics because on its own it is not a valid indicator of high quality patient care. In fact, since the CDC guideline (Dowell, 2016) was issued, there have been many reports of patients who have been successfully managed on opioid analgesics for long periods of time, and in whom the benefits of such therapy exceed the risks, of being forced to abruptly reduce or discontinue their medication regimens. Such involuntary tapers are associated with sometimes extremely adverse outcomes, including depression, anxiety and emergence of other mental health disorder, loss of function and the ability to perform daily activities, and even suicide. There has been considerable discussion of these unintended consequences at meetings of the HHS Interagency Pain Management Best Practices Task Force. In addition, research continues to demonstrate that individuals may or may not have access to pain management therapies based on their race/ethnicity and measures that may further exacerbate this issue should be avoided (Goshal, 2020).

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

Our specific concerns with this measure include the misalignment of the numerator requirements with the evidence and the need for additional precision in the denominator.

Measures that call for hard limits and lead to abrupt tapering or discontinuation of opioids for those already receiving these medications are not consistent with the guideline recommendations (Dowell, 2019). For example, identifying those patients for whom the daily prescribed morphine milligram equivalents (MME) are considered high may serve as an indicator of whether a patient is at risk of overdose and should be co-prescribed naloxone, but it alone is not an appropriate marker of the quality of care provided. The CDC recommendations allow physicians to document a clinical rationale or justification when suggested dose levels are exceeded; yet, the inclusion of an absolute MME requirement does not capture if a justification exists nor does it provide a well-defined and targeted denominator. We have

similar concerns with the inclusion of prescriptions that exceed 90 days as it does not address the needs of those individuals with chronic pain.

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

The measure developer should explore more appropriate methods to assess a patient's chronic pain such as the Pain Assessment Screening Tool and Outcomes Registry (PASTOR) and use this patient-reported data on areas as the basis for performance measures. This tool utilizes the Patient Reported Outcomes Measurement Information System (PROMIS) and through the use of Computer Adaptive Testing, key domains such as sleep disturbance and physical function can be assessed in a targeted and patient-directed way.

In addition, this measure as currently specified lacks the precision needed to ensure that only those patients as defined by the clinical recommendations are included in the denominator. The AMA believes that no measure addressing opioid use should be endorsed and/or used until each is reviewed against the guideline to ensure consistency with its intent. Specifically, the CDC clarified that the guideline is intended to apply to primary care clinicians who treat adult patients for chronic pain (Dowell, 2019). In addition, the CDC stated in a letter to three specialty societies on February 28, 2019 that the recommendations do not apply to those patients receiving active cancer treatment, palliative care, and end-of-life care as well as those with a diagnosis of sickle cell disease (CDC, 2019).

On review of the specifications, the denominator population does not reflect the right population of patients consistent with the evidence. We do not believe that inclusion of some of these conditions within the risk adjustment approach such as individuals with a cancer diagnosis or sickle cell disease is sufficient; rather, these individuals and those receiving palliative care and not just hospice must be excluded.

The measure also lacks the precision needed to ensure that only those patients for whom inappropriate concurrent prescribing of an opioid and benzodiazepine are included in the denominator. Specifically, the patient population could likely include patients for whom concurrent prescribing of these medications may be appropriate, particularly those with chronic pain.

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

Given these significant concerns, the AMA does not support the endorsement of this measure.

***References*:**

AMA letter to CDC re: 2016 Guideline for Prescribing Opioids for Chronic Pain. Dated June 16, 2020. Available at: https://searchlf.ama-assn.org/letter/documentDownload?uri=%2Funstructured%2Fbinary%2Fletter%2FLETTERS%2F2020-6-16-Letter-to-Dowell-re-Opioid-Rx-Guideline.pdf

CDC letter to NCCN, ASCO, and ASH. Dated February 28, 2019. 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

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.

Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain — United States, 2016. MMWR Recomm Rep 2016;65(No. RR-1):1–49. DOI: http://dx.doi.org/10.15585/mmwr.rr6501e1

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

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.

Comment by: The Federation of American Hospitals (FAH)

The Federation of American Hospitals (FAH) and its members actively seek to prevent unintentional opioid overdose fatalities and support measures that address the opioid epidemic but we also believe that any measure in this area must be aligned with current clinical guidelines and its potential unintended consequences must be addressed prior to endorsement.

In response to the misapplication of the recommendations from the Centers for Disease Control and Prevention (CDC) Guideline for Prescribing Opioids for Chronic Pain — United States, 2016, the guideline authors published an article in the New England Journal of Medicine seeking to clarify its intent and are also in the process of updating the guidelines to address some of these issues (Dowell 2016, Dowell 2019). Specifically, the authors were concerned that these discrepancies could potentially lead to patient harms through abrupt tapering or discontinuation of opioids for current users of high opioid dosages and/or inclusion of patient populations for whom chronic use or higher dosages may be warranted. Based on the FAH's comparison of this measure against the CDC guideline recommendations, we believe that it is not currently supported by the recommendations.

Specifically, the intent of the CDC guideline was to address the care provided by primary care providers for patients with chronic pain and the current population captured in the measure is not aligned with the evidence. For example, the measure is likely to include patients who are already receiving both an opioid and a benzodiazepine or opioids that exceed the morphine milligram equivalents threshold or the 90-day timeframe. The FAH does not believe that there is strong evidence to support abrupt discontinuation of these therapies, instead tapering should be considered. Requiring that these drugs be discontinued to meet performance on a measure alone is not appropriate and has the potential to compromise patient safety and lead to patient harm.

In addition, the patient population must be further narrowed to capture the additional diagnoses where it is appropriate to use these medications including those with sickle cell disease, active cancer, and palliative care. These additional exclusions are supported in the NEJM article as they explicitly state that the recommendations do not apply to these populations. While we note that some of the clinical variables for these diagnoses are included in the risk adjustment approach, the FAH believes that it would be more appropriate to exclude these populations from the measure.

This measure could result in providers not offering suitable pain solutions to patients receiving dialysis, which is contrary to the goal of a positive patient care experience if these treatments are needed. Reframing this measure to focus on adequate pain assessments and treatments would assist all of us in understanding the true problem rather than removing a downstream intervention.

Thank you for the opportunity to comment.

References:

Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain — United States, 2016. MMWR Recomm Rep 2016;65(No. RR-1):1–49. DOI: http://dx.doi.org/10.15585/mmwr.rr6501e1

Dowell D, Haegerich T, Chou R. No shortcuts to safer opioid prescribing. N Engl J Med. 2019 Apr 24. doi: 10.1056/NEJMp1904190. [Epub ahead of print]

Comment by: Kidney Care Partners (KCP)

Kidney Care Partners (KCP) appreciates the opportunity to submit early (pre-Standing Committee meeting) comments on the measures under consideration for endorsement in the National Quality Forum's Renal Project Spring 2021 Cycle. KCP is a coalition of members of the kidney care community that includes the full spectrum of stakeholders related to dialysis care—patient advocates, healthcare professionals, dialysis providers, researchers, and manufacturers and suppliers—organized to advance policies that improve the quality of care for individuals with both chronic kidney disease and end stage renal disease. We commend NQF for undertaking this important work. The following comments apply to both measures under review this cycle:

NQF 3615: Unsafe Opioid Prescriptions at the Dialysis Prescriber Group Level (CMS)

NQF 3616: Unsafe Opioid Prescriptions at the Dialysis Practitioner Group Level (CMS)

Overarching Comments

KCP recognizes the profound importance of minimizing opioid overuse in dialysis patients and appreciates the underlying intent of these measures; however, we have serious concerns with both as currently specified and cannot offer our support of either. Recognizing that opioids have been overused previously, it is important to note that national efforts have resulted in a substantial decrease in prescription opioid use in the past several years. Based on CDC data, prescription opioid dispensing rate in 2019 was 57% of the peak in 2012, and these data do not account for the changes in prescribing patterns that also have resulted in fewer opioids being dispensed per prescription in recent years. Critically, there are many reasons for extended use of opioids in the dialysis population, where the burden of symptoms is extremely high, life expectancy in many patients is half that in the age-similar general population, and options for pain medications are limited due to safety factors with other agents—for example, gabapentin and pregabalin may have serious neurologic consequences in dialysis patients, while non-steroidal anti-inflammatory drugs may be contraindicated in many individuals with ESRD (e.g., those with residual kidney function and at heightened bleeding risk). These factors question the assertion in the name of the proposed metrics that all opioid use for more than 90 days is 'unsafe.' KCP believes these proposed metrics will incentivize inappropriately abrupt reductions of opioid medications and undermanagement of chronic pain in complex dialysis patients, particularly in the absence of existing knowledge on how to reduce opioid use while sufficiently treating pain in the hemodialysis population. We also believe the measures as specified will exacerbate existing sociodemographic, economic, and geographic disparities related to opioid use, and will result in untenable and specious double penalties for many nephrology groups. Finally, we highlight critical ongoing research from the NIH in the hemodialysis population evaluating patient-centered strategies for promoting safe and durable opioid use reduction while adequately managing pain (HOPE Consortium Trial to Reduce Pain and Opioid Use in Hemodialysis, NCT04571619).

The history of pain management in the United States is complex, oscillating between extremes. While in the midst of an unprecedented opioid epidemic, it is easy to lose sight of our past. Millions of Americans with advanced and debilitating disease suffered needlessly in the 1980s because physicians were overly cautious about prescribing narcotics. We fear these measures portend a return to such days and will ultimately do more harm than good.

Our specific concerns with the measures follow.

Potential for Unintended Consequences is Substantial

We note that, pursuant to the 2018 SUPPORT (Substance Use-Disorder Prevention that Promotes Opioid Recovery and Treatment) Act, HHS contracted with the National Quality Forum (NQF) to convene a Technical Expert Panel (TEP) to review quality measures related to opioids. In its February 2020 report, that TEP explicitly recommended opioid measures to be used in Federal quality programs should address any of a number of patient-centric clinical issues, such as recovery from opioid use disorder (OUD),

assessment and treatment of physical and mental health comorbidities to OUD, co-prescription of naloxone, patient-centered analgesia, and appropriate opioid tapering. The two proposed opioid safety measures address none of those topics, instead focusing exclusively on reducing opioid use—without regard for clinical decision-making or consideration of the etiology or severity of the pain, or the impact on the patient's quality of life.

While the research by Kimmel et al,1 cited as evidence supporting both measures, did find an association between opioid prescription and death, dialysis discontinuation, and hospitalization in dialysis patients, the authors make clear that an opioid prescription may merely be a marker of more severe or advanced illness in dialysis patients and that a causal relationship with these adverse outcomes cannot be inferred. Importantly, Kimmel also referred to evidence that pain is pervasive in individuals with ESRD2,3,4,5 and is linked to a significantly diminished quality of life,6,7,8,9 and that while aggressive pain treatment has been advocated,10,11,12 underestimation and undertreatment of pain still occur in dialysis patients.13,14 These truths are not taken into consideration in these measures.

We note that the NIH-sponsored Hemodialysis Opioid Prescription Effort (HOPE) Consortium (NCT04571619), shepherded by Dr. Kimmel, is actively researching pain and opioid use in the ESRD population and how to safely decrease dependence in dialysis patients, including such behavioral/ cognitive interventions as pain coping skills and use of medications such as buprenorphine. This research aims to develop personalized treatments based on individual patient needs—a critical consideration, given the varied and notoriously persistent nature of pain in this complex and vulnerable population.

Understanding the epidemiology of pain in patients on dialysis—as well as patients' unique needs and preferences—is crucial for further improvement in managing pain. These proposed measures clearly miss that mark. We believe the development of more appropriate measures may be feasible once findings from the HOPE Study are disseminated and digested. Adoption of measures addressing such a crucial aspect of care prematurely, absent this critical knowledge, will do little to improve dialysis care or patient outcomes; rather, we fear these performance measures may induce a range of unintended, deleterious, and potentially profound adverse consequences.

Double Penalties

From the specifications and supporting measure information, it appears that the attributable entity for the Practitioner Measure is the treating nephrologist's group practice, irrespective of who prescribed the opioid—whether the nephrologist herself or a physician entirely unrelated to her group. The nephrologist is thus held accountable for other providers' prescriptions. Additionally, as the attributable entity with the Prescriber Measure is the opioid prescriber, implementation of both measures together in a payment program would seemingly result in nephrology groups being penalized twice when the nephrologist is also the opioid prescriber. We see no indication in the measure materials that this would not be the case.

Sociodemographic and Geographic Disparities

Finally, while unsafe opioid use was found to be associated with White race, non-Hispanic ethnicity, dual eligible status, and unemployment in UM-KECC's analyses, gender was the only SDS/SES factor15 included in the final risk models because "… it is unclear whether [these] associations… are due to underlying biological or other patient factors or represent disparities in care. Adjusting for these social risk factors could have the unintended consequence of creating or reinforcing disparities and facilitating unsafe prescribing practices." As KCP has commented in the past (see, for example, KCP's August 2018 QIP comment letter to CMS), we agree CMS must strike the correct balance to ensure that it meets the goals of both fairly assessing providers while also not masking potential disparities or disincentivizing the provision of care to more medically complex patients. However, we reiterate our strong preference for adopting an SDS adjustment for measures where it has been shown that SDS factors are driving differences in the outcomes being reported. Given the associations noted above, KCP believes gender as the only sociodemographic risk variable is insufficient and is concerned the measures risk potentiating

existing health inequities. We believe other biological and demographic variables are important, and not accounting for them is a significant threat to the validity of both measures.

In a similar vein, Kimmel et al [2017] reported geographic trends in opioid use in patients with ESRD are comparable to those in the general population, with eight states having chronic opioid prescription rates of 30% or more. "Chronic opioid prescription rates ranged from 9.5% of patients on dialysis in Hawaii to 40.6% of patients in West Virginia in 2010. Seven other states had prescription rates >30% (Michigan, Oklahoma, Oregon, Kentucky, Idaho, Indiana, and Alabama):"16

Yet it does not appear from the supplied risk model data that geography itself (distinct from the Area Deprivation Index) was examined. The failure to do so when such regional variations in opioid use is well-documented is puzzling, at best.

Given these empirically demonstrated sociodemographic and geographic opioid use disparities, KCP is not convinced that these measures have been sufficiently adjusted to avoid exacerbating existing inequities, disincentivizing the provision of care to more medically complex patients, and adversely impacting quality of life for our most vulnerable patients.

Technical Concerns

In addition to our above core conceptual issues, we also note the following technical concerns with the measures:

Patient Exclusions. Again, KCP is concerned that the measures as specified may result in the under-treatment of pain in patients in whom longer-term use of opioids is warranted. As such, we believe the single patient-level exclusion for hospice is insufficient in measures addressing opioid use, overlooking the many patients suffering with debilitating chronic pain (even unrelated to ESRD) and those with a life-threatening comorbidity not yet eligible for hospice care. Notably, this metric again highlights the real-world limitations in accessing hospice services among patients receiving maintenance hemodialysis. We believe additional exclusions for patients with claims for palliative care and for those under the care of a pain management specialist during the reporting period would strengthen the measure considerably.

Reliability—Profile Inter-Unit Reliability (PIUR). KCP has consistently opposed CMS's use of the PIUR for accountability metrics intended to distinguish performance between providers. CMS crafted this novel metric of reliability to "assess more directly the value of performance measures in identifying facilities with extreme outcomes."17 Per CMS: "The PIUR indicates the presence of outliers or heavier tails among the providers, which is not captured in the IUR itself. . . . [When] there are outlier providers, even measures with a low IUR can have a relatively high PIUR and can be very useful for identifying extreme providers." KCP strongly concurs, however, with NQF's Scientific Methods Panel (SMP) that the PIUR is not an appropriate reliability metric for measures in any accountability program intended to distinguish performance between providers falling in the middle of the curve, along a continuum. The ability to reliably distinguish outliers is inconsistent with the purpose of such programs, and the SMP concluded the IUR is and remains the appropriate reliability statistic for this purpose. While in this instance the measures' IURs are acceptable, KCP on principle reiterates its general opposition to use of the PIUR to demonstrate reliability in accountability metrics used in programs intended to distinguish performance along a curve.

Validity: Validity was tested at the performance measure scores by evaluating the concordance between the measure scores, hospitalization metrics, and mortality rates. With mortality, to account for potential selection bias stemming from the fact that the definition of chronic opioid use requires patients survive at least 90 days (e.g., those who survived 90+ days may be healthier), patients were instead stratified based on length of time at risk during the 12-month performance period. It is not clear to us, however, how the ensuing time at risk stratification was performed, and we are unable to replicate the results with the information provided. We also note that p-values were not included for the mortality stratification and we thus cannot confirm the results are statistically significant. We request clarification on UM-KECC's approach to these calculations, accompanied by an appropriate assessment of significance to allow for a thorough assessment of the measures' validity.

Another essential component of measure validity is demonstration of meaningful differences in performance, allowing end-users of public reporting or value-based purchasing programs to make informed decisions about the quality of care delivered by various providers. Here, for each provider group the proportion of patient-months with a high-risk opioid prescription was calculated at the year-level and then was compared to the overall national distribution, yielding the following results:

Practitioner Groups

Better than Expected - 122 (3.67%)

As Expected - 3,092 (93.05%)

Worse than Expected - 109 (3.28%)

Prescriber Groups

Better than Expected - 309 (6.03%)

As Expected - 4,635 (90.47%)

Worse than Expected - 179 (3.49%)

While UM-KECC concludes its analysis demonstrates both practical and statistically significant differences in performance, it should be noted that the measures only distinguish performance in <7% and <10% of practitioner and prescriber groups, respectively, with the overwhelming majority of measured entities performing "as expected." A performance measure in which greater than 90% of all measured entities are reported as performing "as expected" provides little meaningful, actionable information to patients, and we are not convinced these statistics are sufficiently compelling to support the measures' use in publicly reported accountability programs.

Risk Model*:* In prior comments to UM-KECC and CMS on measures with similar risk models, KCP has noted that many of the prevalent comorbidities in the final model have p-values significantly greater than 0.05 (e.g., prostate and renal cancer, headaches, osteomyelitis). While in the past CMS/UM-KECC has responded that the large number of clinical factors in such models generates multicollinearity among covariates, likely resulting in some unexpected results, we remain concerned that this strategy results in a model that will not be generalizable. In the opioid models, for example, allergic reactions are associated with a higher risk of unsafe opioid use than breast or peritoneal cancers. While KCP has consistently voiced its support of prevalent comorbidity adjustment, we have in the past posited that these illogical findings are a function of collinearity and coding idiosyncrasies that may result in the proposed collection of adjusters becoming less robust with each year that passes from initial model development.

KCP also notes that validity testing yielded c-statistics of 0.70 and 0.74 for the practitioner and prescriber measures, respectively. We are concerned the model will not adequately discriminate

performance—particularly that smaller units might look worse than reality. We believe a minimum c-6

statistic of 0.8 is a more appropriate indicator of the model's goodness of fit and validity to represent meaningful differences among facilities and encourage continuous improvement of the model.

KCP again thanks you for the opportunity to comment on this important work. If you have any questions, please do not hesitate to contact Lisa McGonigal MD, MPH (lmcgon@msn.com or 203.539.9524).

Sincerely,

Kidney Care Partners

References:

1 Kimmel PL et al. Opioid prescription, morbidity, and mortality in United States Dialysis Patients. JASN. 2017;28(12):3658-3670.

2 Raghavan D, Holley JL. Conservative care of the elderly CKD patient: A practical guide. Adv Chronic Kidney Dis. 2016;23:51–56.

3 Davison SN. Pain in hemodialysis patients: Prevalence, cause, severity, and management. AJKD. 2003;42:1239–1247.

4 Santoro D et al. Pain in end-stage renal disease: A frequent and neglected clinical problem. Clin Nephrol. 2013;79[Suppl 1]:S2–S11.

5 Shayamsunder AK et al. Sleepiness, sleeplessness, and pain in end-stage renal disease: Distressing symptoms for patients. Semin Dial. 2005;18:109–118.

6 Davison SN. Pain in hemodialysis patients: Prevalence, cause, severity, and management. AJKD. 2003;42:1239–1247.

7 Harris TJ et al. Pain, sleep disturbance and survival in hemodialysis patients. Nephrol Dial Transplant. 2012;27:758–765.

8 Davison SN. Chronic kidney disease: Psychosocial impact of chronic pain. Geriatrics. 2007;62:17–23.

9 Davison SN, Jhangri GS. Impact of pain and symptom burden on the health-related quality of life of hemodialysis patients. J Pain Symptom Manage. 2010;39:477–485.

10 Barakzoy AS, Moss AH. Efficacy of the world health organization analgesic ladder to treat pain in end-stage renal disease. JASN. 2006;17:3198–3203.

11 Claxton RN et al. Undertreatment of symptoms in patients on maintenance hemodialysis. J Pain Symptom Manage. 2010;39:211–218.

12 Davison SN, Koncicki H, Brennan F. Pain in chronic kidney disease: A scoping review. Semin Dial. 2014;27:188–204.

13 Barakzoy AS, Moss AH. Efficacy of the world health organization analgesic ladder to treat pain in end-stage renal disease. JASN. 2006;17:3198–3203.

14 Merboth MK, Barnason S. Managing pain: The fifth vital sign. Nurs Clin North Am. 2000;35:375–383.

15 Per CMS, biologic differences (e.g., genetic, hormonal, metabolic) may account for differences in pain perception between male and female, suggesting a physiologic effect rather than a disparity in care.

16 Kimmel PL et al. Opioid prescription, morbidity, and mortality in United States Dialysis Patients. JASN. 2017;28(12):3658-3670.

17 Kalbfleisch JD, He K, Xia L, Li Y. Does the inter-unit reliability (IUR) measure reliability? Health Services and Outcomes Research Methodology. 2018;18(3):215-225. Doi: 10.1007/s10742-018-0185-4.

Of the 2 NQF members who have submitted a support/non-support choice:

1 supports the measure

1 does not support the measure

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3615

Measure Title: Unsafe Opioid Prescriptions at the Prescriber Group Level

RELIABILITY: SPECIFICATIONS

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

Submission document: "MIF_3615" 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.*

Briefly summarize any concerns about the measure specifications.

Panel Member 2: Minor: As noted below, it wasn't clear to me whether prevalent comorbidities in the risk adjustment model were ascertained from claims submitted during the measurement period or some prior period.

Panel Member 3: The attribution of patients to groups is not well described. The impact of "double counting" of patients receiving prescriptions from more than one practice is also not addressed.

Panel Member 4: Two issues regarding measure specifications: [1] The MIF (in S.7) notes the denominator includes "dialysis patient[s]". However, this is not defined here nor in the XL data dictionary file. [2] The MIF (in S.8) notes an exclusion of "Patients who have a hospice claim". In S.9 we only receive a high-level definition of hospice, but it's definitely not clear how this is exclusion is defined based on this response. Example: Unstated as to what "CMS file" is referenced. Unstated what fields in said form & what responses / codes are employed to operationally meet the hospice definition so as to exclude a given case.

Panel Member 6: Question #1 - Please note - that I do not understand question #1 above so my answer is arbitrary. Question #2 - No major concerns

Panel Member 7: The developers did not include value sets for opioids or benzodiazepines. I am also unclear about the duration of the reporting period, as it is not clearly stated anywhere (presumably one year). Finally, it is unclear what happens when multiple prescriptions from multiple prescribers collectively contribute >90 days; do all prescribers get blamed, even the first in the sequence of prescriptions? The one-month minimum duration of Part D enrollment is not consistent with the duration component of the numerator.

Panel Member 9: I think the specifications are accurate and precise, but even after several reads of the developer's definitions, I'm struggling with the distinction between prescriber group and provider group.

RELIABILITY: TESTING

Type of measure:

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

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

Data Source:

☐ **Abstracted from Paper Records** ☒ **Claims** ☒ **Registry** ☐ **Abstracted from Electronic Health Record (EHR)** ☐ **eMeasure (HQMF) implemented in EHRs** ☐ **Instrument-Based Data** ☒ **Enrollment Data** ☒ **Other (please specify)**
Panel Member 4: IDR Medicare Provider table selected for MCPs
Panel Member 5: IDR Medicare Provider table selected for MCPs
Panel Member 7: The developers claim to use registry data, but it is not clear what they mean by a registry. Is CROWNWeb considered a registry?

Level of Analysis:

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

Measure is:

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

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

- **Reliability testing level ☒ Measure score ☐ Data element ☐ Neither**
- **Reliability testing was conducted with the data source and level of analysis indicated for this measure ☒ Yes ☐ No**
- 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**
- **Assess the method(s) used for reliability testing**
- **Submission document:** Testing attachment, section 2a2.2
- **Panel Member 1:** IUR and PIUR
- **Panel Member 3:** IUR and PIUR were used to estimate provider level reliability, and outlier effects for groups with at least 11 patients. The split sample analysis as described states that patients were divided into two equal groups within practice, and that the process was repeated 100 times. For small groups that process would yield uninterpretable results. The average number of patients per group was 40 and the average number of physicians per group was 20. It is not clear how the variable number of physicians per group was handled in this process.
- **Panel Member 4:** The use of ANOVA testing in this circumstance is appropriate. No concerns.
- **Panel Member 5:** IUR: The overall IUR is 0.86, which means 86% of the total variation of this prescriber group level measure can be explained by the differences among prescribers and not by random noise.
- **Panel Member 6:** The method is appropriate.
- **Panel Member 7:** Inter-unit reliability and profile IUR were estimated using a one-way ANOVA approach with 100 resampled datasets.
- **Panel Member 8:** IUR and PIUR for providers with 11 or more eligible patients
- **Panel Member 9:** Methods were appropriate
- **Assess the results of reliability testing**
- **Submission document:** Testing attachment, section 2a2.3
- **Panel Member 1:** IUR = 0.86 PIUR = 0.96
- **Panel Member 3:** The variation between providers within provider group does not appear to have been handled by the methods reported (i.e., the error term appears not to include between providers across patients within practice).
- **Panel Member 4:** The testing results were an IUR of 0.86 and a PIUR of 0.98. The difference in the results is accounted for. The testing results demonstrates the measure is reliable.
- **Panel Member 5:** IUR: The overall IUR is 0.86, which means 86% of the total variation of this prescriber group level measure can be explained by the differences among prescribers and not by random noise. Reliability is high. However, with a value this high, it is possible that the risk adjustment is not adequate.
- **Panel Member 6:** The test sample is adequate to generalize.
- **Panel Member 7:** Overall IUR=0.86, very good. No comment on how this may vary according to the "size" of the provider group.
- **Panel Member 8:** IUR 0.86; PIUR 0.98
- **Panel Member 9:** Very strong IUR and PIUR values are reported.

- 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)
- 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)
- **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)
- **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**
- **Panel Member 1:** Very high reliability
- **Panel Member 2:** Estimated reliability = 0.86
- **Panel Member 3:** The variation between providers within provider group does not appear to have been handled by the methods reported (i.e., the error term appears not to include between providers across patients within practice).
- **Panel Member 4:** Two issues regarding measure specifications: [1] The MIF (in S.7) notes the denominator includes "dialysis patient[s]". However, this is not defined here nor in the XL data dictionary file. [2] The MIF (in S.8) notes an exclusion of "Patients who have a hospice claim". In S.9 we only receive a high-level definition of hospice, but it's definitely not clear how this is exclusion is defined based on this response. Example: Unstated as to what "CMS file" is referenced. Unstated what fields in said form & what responses / codes are employed to operationally meet the hospice definition so as to exclude a given case.
- **Panel Member 5:** IUR: The overall IUR is 0.86, which means 86% of the total variation of this prescriber group level measure can be explained by the differences among prescribers and not by random noise. Reliability is high. However, with a value this high, it is possible that the risk adjustment is not adequate.
- **Panel Member 6:** No major concerns.
- **Panel Member 8:** with such a wide variation in the number of patients per provider it is surprising that reliability so high so perhaps I am missing something.
- **Panel Member 9:** The IUR and PIUR results are high enough to rate as "high", assuming that the calculations are technically correct.

VALIDITY: TESTING

1. **Validity testing level: ☒ Measure score ☐ Data element ☐ Both**
2. **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)

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

   ☐ **Face validity**

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

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

4. **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)

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

   Submission document: Testing attachment, section 2b2.2

   Panel Member 1: Associations with mortality and hospitalization rates

   Panel Member 2: For me, the main validity questions pertain to the choice of numerator criteria and case mix adjustment. I would like to know that there is ample evidence/consensus that numerator occurrences necessarily reflect a lapse in care quality (as opposed to rare cases where such prescribing is actually appropriate) and that the measure adequately adjusts for potential differences in pain severity.

   Panel Member 3: The hospitalization rates, days of hospitalization and mortality rates appear to be estimated at the practice/clinician level. However, the opioid use rate appears to be only estimated at the patient level. The measure is to be used at the practice level.

   Panel Member 4: Assessing the correlation between 3615 and hospitalization and mortality is an appropriate validity test given the provided discussion of the literature. However, there is not specific correlation test specified. It appears the relationships are stated with descriptive statistics.

   Panel Member 6: The method seems reasonable.

   Panel Member 7: Developers estimated differences across tertiles of performance in patient hospitalization rates and mortality rates. This is a promising approach, but the observed associations could be explained entirely by case mix/severity. Did the developers use risk-adjusted or unadjusted performance scores? Did they risk-adjust their hospitalization and mortality rates? Did they demonstrate any specificity to these relationships - i.e., hospitalizations or deaths potentially linked to opioid use? Did they consider the potentially adverse effects of abruptly stopping opioids at 90 days or less?

   Panel Member 8: correlation with mortality and hospitalization rates--sound clinical basis for validity testing

   Panel Member 9: Correlations with other quality measures would seem to establish validity in this particular example. The evidence is not strong and compelling, but it is not zero either.

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

Submission document: Testing attachment, section 2b2.3

Panel Member 1: Weak association with hospitalization, as expected Stratified by months at risk, unsafe opioid use more strongly associated with mortality risk at patient-level, but not at practice-level

Panel Member 2: The results are supportive of validity

Panel Member 3: Variation in hospitalization rates and mortality rates between groups is very small for tertiles of opioid use, despite significant trend results in the first case. Comparisons of safe vs. unsafe opioid use were apparently performed at the patient-level for months at risk results, were highly variable across time intervals, and the absolute risk increase was 2.87%. It is not clear that the analyses considered the magnitude of "unsafe use."

Panel Member 4: The relationships between 3615 and hospitalization and mortality appear to be low to modest. Ascertaining a degree of correlation is challenging as there is not a specific correlation test specified. Results are displayed and discussed primarily in terms of descriptive statistics.

Panel Member 5: assessed the association between unsafe opioid use and mortality, stratified by the number of months at risk. Within each stratum, unsafe opioid use is associated with higher mortality.

Panel Member 6: The test sample is adequate to generalize to widespread implementation.

Panel Member 7: The expected associations were supported but seem (to me) likely due to confounders (case mix), given that long term NH residents and cancer patients are more likely to have "unsafe opioid prescriptions." Graphical representation of the relationships, or an analysis of correlation across the entire distribution of the measures, would be helpful.

Panel Member 8: strong association with both mortality and hospitalization

Panel Member 9: Reported correlations are statistically significant and in the predicted direction.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

Submission document: Testing attachment, section 2b2.

Panel Member 4: No concerns.

Panel Member 6: The only exclusion noted is hospice. This seems reasonable.

Panel Member 7: none

Panel Member 9: None

8. **Risk Adjustment**

Submission Document: Testing attachment, section 2b3

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

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

☐ Yes ☐ No ☒ Not applicable

19c. **Social risk adjustment:**

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

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

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

19d. **Risk adjustment summary:**

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

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

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

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

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

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

Panel Member 1: Model with demographic factors and comorbid conditions

Panel Member 2: The developers discuss the rationale for various modeling choices, but they didn't say much about the decision to adjust in the first place. I think it makes sense, but I could also imagine arguments against adjustment. The model adjusts for age, sex, BMI, duration of ESRD, nursing home status in previous year, diabetes as primary cause of ESRD, comorbidities at ESRD incidence, and prevalent comorbidities. It wasn't clear to me whether prevalent comorbidities are assessed based on claims submitted during the measurement period or during a time interval preceding the measurement period. If the former, then strictly speaking model is violating the principle that adjustment variables should be present before the start of care. This concern is probably more theoretical than practical. The model appears to be well calibrated overall. I was curious to know if the developers assessed calibration within subgroups such as patients with multiple comorbidities. Assessing calibration by race could shed light on the potential consequences of not adjusting for race despite the apparent strong odds ratios for race categories. The calibration plot Figure 4 compares observed versus expected based on absolute numbers of patients instead of proportions (probabilities). I was curious to know if calibration looked equally good when plotting proportions.

Panel Member 3: Although SDS/SES were statistically significant in risk adjustment models, the only variable retained was gender based on the movement across "better," "the same" and "worse" than expected performance categories of < 1% of facilities with and without SDS/SES variables.

Panel Member 4: The risk adjustment strategy is well constructed. The method to identify and test covariates is appropriate. All risk factors were present at the onset of the measurement period. Social risk factors were identified, tested for inclusion, and an analysis was performed as to whether / the degree to which they impact the ratings.

Panel Member 5: Regarding question 19d.4: cannot evaluate because model was not validated in a validation data set.

Panel Member 6: I agree with the rationale of the developer.

Panel Member 7: Risk-adjustment approach is very poorly justified. Process measures are rarely risk-adjusted because the presumption is that "safe and effective" care is linked to a specific denominator-eligible population (after exclusions and stratification, as appropriate). A complex risk-adjustment model of this type belies the concept of "unsafe opioid prescriptions." Implicit in their risk-adjustment approach is the concept that "unsafe prescribing" must be safe and effective for some subsets of the eligible population, or else why would we account for these patient characteristics in a process measure? Their models include age, sex, BMI, time on ESRD, nursing home residence, cause of ESRD, and hundreds of comorbid conditions. Some of these covariates, like age, make clinical sense (i.e., it is clearly safer to prescribe medium-dose opioids to younger patients than to older patients). But how do the duration of ESRD or long-term NH residence, for example, "justify" so-called "unsafe opioid prescriptions"? Even worse, the list of covariates includes features that don't make clinical sense in this context, such as "renal failure," "chronic kidney disease," and "drug dependence."

Panel Member 8: very comprehensive list of variables and appropriate model development. c-statistic 0.74 and reasonable assessment of impact of excluding SR factors (only one group changed status)

Panel Member 9: The developers have done a careful job of analyzing both clinical and social risk factors; they have decided to retain a large number of clinical variables with both significant and non-significant effects and have chosen to leave out social variables with essentially the same magnitude of effect. They

then show that the results for entities being compared are not different with social factors included or not included. The same analyses and same observations could have been done with any similar set of clinical variables, but the developers have not chosen that approach to analysis. This violates the 2014 NQF Expert Panel recommendation that clinical and social variables be treated the same, but most developers still do not treat clinical and social variables the same, so it would be arbitrary to fail this measure on this argument.

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

   Submission document: Testing attachment, section 2b4.

   Panel Member 2: The developers did not present the overall rate of unsafe prescribing or estimates of signal variation these rates across providers. Various graphs and tables showed observed rates, but these are impacted by random statistical variation and therefore over-estimate the true signal variation. Based on the high reliability estimates I assume that the signal variation is quite large and hence the differences are very meaningful. The developers mention that the reporting of the measure is limited to providers with at least 11 eligible cases. This seems like a small number. I wonder if the developers could estimate what reliability would be if the measure was estimated using a random sample of 11 patients per provider. I wonder how many providers would be excluded if the threshold was increased to say 20?

   Panel Member 3: The proportion of provider groups with statistically significant differences in performance (better or worse than expected) was small, with only 3.5% of provider groups having unsafe opioid practices. Given measurement error, it is not clear whether prescription misclassification bias would further reduce the proportion of practices in the unsafe group.

   Panel Member 4: No concerns. There is a fair degree of outliers via statistically tested results. More specifically, 6% of providers were classified as "better", and 3.5% were "worse".

   Panel Member 6: No major concerns.

   Panel Member 8: appears to identify outliers

   Panel Member 9: There is no way to tell how big a difference is meaningful.

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

    Submission document: Testing attachment, section 2b5.

    Panel Member 4: NA – 1 set of specifications was used for this measure.

    Panel Member 6: No major concerns.

    Panel Member 7: Not applicable

    Panel Member 8: linkage of data sources likely appropriate

    Panel Member 9: N/A

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

    Submission document: Testing attachment, section 2b6.

    Panel Member 3: The developer did not report the overall rate of missing data, only that for risk adjustment variables.

    Panel Member 4: Surprising that for cases with missing BMI that the case is assigned the highest level of risk for BMI (i.e., 30+). That method has the unintended consequence to game the measure by omitting BMI. This potential for gaming is likely the reason that for other measures with missing data that the lowest level of risk is assigned (where the missing data is used in risk adjustment). The result is artificially driving up the risk adjustment. However, the impact on overall risk adjusted rate is likely minor given the coefficients for the BMI variable.

    Panel Member 6: No major concerns.

Panel Member 7: Addressed adequately.

Panel Member 8: minimal missing data--unlikely to introduce bias

Panel Member 9: None

For cost/resource use measures ONLY:

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

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

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

14. **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.)

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

    Panel Member 1: c = 0.74 and good calibration

    Panel Member 3: Variation in hospitalization rates and mortality rates between groups is very small for tertiles of opioid use, despite significant trend results in the first case. Comparisons of safe vs. unsafe opioid use were apparently performed at the patient-level for months at risk results, were highly variable across time intervals, and the absolute risk increase was 2.87%. It is not clear that the analyses considered the magnitude of "unsafe use."

    Panel Member 4: The relationships between 3615 and hospitalization and mortality appear low to modest. Ascertaining a degree of correlation is challenging as there is not a specific correlation test specified. Results are displayed and discussed in terms of descriptive statistics.

    Panel Member 5: there is insufficient data to determine the validity of the risk adjustment model because model performance was not evaluated in a validation data set.

    Panel Member 7: See above, the validation approach is probably confounded by case mix (although it is unclear) and the risk-adjustment approach is poorly justified.

    Panel Member 8: Appropriate risk adjustment and validity testing

    Panel Member 9: Correlational analysis supports validity of the measure.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

☐ **Insufficient**

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

ADDITIONAL RECOMMENDATIONS

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

    Panel Member 7: When is it appropriate to risk-adjust process measures, such as measures of "unsafe" prescribing? This term implies - to me - that the goal is zero, and thus risk-adjustment is inappropriate. But the any-or-none design of this measure actually includes a lot of perfectly safe prescribing, which is why they need to risk-adjust. The measurement concept is muddled.

Brief Measure Information

NQF #: 3615

Corresponding Measures:

De.2. Measure Title: Unsafe Opioid Prescriptions at the Prescriber Group Level

Co.1.1. Measure Steward: Centers for Medicare and Medicaid Services

De.3. Brief Description of Measure: Percentage of all dialysis patients attributable to an opioid prescriber's group practice who had an opioid prescription written during the year that met one or more of the following criteria: duration >90 days, Morphine Milligram Equivalents (MME) >50, or overlapping prescription with a benzodiazepine.

Please note that the opioid prescriber is the clinician identified from Part D Medicare Claims who actually provides an opioid prescription to a dialysis patient. This provider is usually not the nephrologist who is overseeing the patient's dialysis care. This is in contrast to NQF submitted measure #3616, which is at the dialysis provider level (the clinician who receives the Monthly Capitated Payment for overseeing dialysis care). While the dialysis provider is usually not the clinician who is prescribing opioids, the MCP physician does have a responsibility to be aware of dialysis patients medications and that doses are safe and appropriate for level of kidney function.

The proposed measure is a directly standardized percentage, which is adjusted to the national distribution of covariates (e.g., age, gender, risk factors). Here, "national" refers to all opioid prescriber groups combined. Specifically, the standardized rate for a given prescriber's group is an estimate of the group's percentage of unsafe opioid prescriptions if their case-mix were equal to that of the national population. Case-mix adjustment is based on a logistic regression model.

1b.1. Developer Rationale: Several observational studies have demonstrated an association between unsafe opioid use in the dialysis population and higher risk of fall/fracture, hospitalization, and mortality. Unsafe opioid use is typically defined as >50 morphine milligram equivalents (MME), duration > 90 days, or co-prescription with a benzodiazepine.

The measure focus is the process determining the percentage of all dialysis patients attributable to an opioid prescriber's group practice who had an unsafe opioid prescription written within the year.

The measure is risk adjusted to mitigate against the unintended consequences of under treatment of pain in patients with comorbidities that have a significant pain component (e.g., cancer, sickle cell disease). By adjusting for case-mix at the prescriber's group practice, our intent is for providers to be able to write necessary opioid prescriptions for patients with greater comorbidity, and likely greater analgesia needs, since the measure does not penalize individual prescribing events or hold providers to an absolute scale or threshold. Rather, the measure identifies the small number of group practices that, based on their year-long prescribing patterns, have extreme deviations relative to their peers.

Once implemented practitioner performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in reducing unsafe opioid use, while accounting for patients where higher dose or longer term therapy may be warranted.

S.4. Numerator Statement: The numerator is the number of patients in the denominator who were prescribed an opioid that was either >90 days duration during the year, >50 MME, or overlapped in time with a benzodiazepine prescription.

S.6. Denominator Statement: The denominator is the number of patients associated with an opioid prescriber's group practice who are receiving maintenance dialysis (in-center or home dialysis) for any duration who receive an opioid prescription during the one year reporting period.

S.8. Denominator Exclusions: Patients who have a hospice claim at any time (either before or after the opioid prescription date) during the one year reporting period are excluded.

De.1. Measure Type: Process

S.17. Data Source: Claims, Other, Registry Data

S.20. Level of Analysis: Clinician: Group/Practice

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

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

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

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

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

Prescriber_Group_Evidence.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*)**:**

Measure Title: Unsafe Opioid Prescriptions at the Prescriber Group Level

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

Date of Submission: **4/2/2021**

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

Outcome

☐ Outcome:

☐ Patient-reported outcome (PRO):

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

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

☒ Process: Percentage of all dialysis patients attributable to an opioid prescriber's group practice who had an unsafe opioid prescription written

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

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

Several observational studies have demonstrated an association between unsafe opioid use in the dialysis population and higher risk of fall/fracture, hospitalization, and mortality. Unsafe opioid use is typically defined as >50 morphine milligram equivalents (MME), duration > 90 days, or co-prescription with a benzodiazepine.

The measure focus is the process determining the percentage of all dialysis patients attributable to an opioid prescriber's group practice who had an unsafe opioid prescription written within the year.

This process leads to improvement in fractures, hospitalizations, and mortality as follows:

Measure percentage of patients with unsafe opioid prescriptions Ò Assess value Ò Identify patients who have an unsafe opioid prescription Ò Evaluate/change pain management (decrease dose, consider alternative agent, avoid co-prescription with benzodiazepine) Ò lower percentage of unsafe opioid prescription Ò Lower patient fractures, hospitalizations, and mortality.

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*)

X Other

1a.4 OTHER SOURCE OF EVIDENCE

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

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

Pain is among the most commonly reported symptom of patients on dialysis and patients with end stage renal disease (ESRD) report more pain than those in the general population. ESRD patients may be especially vulnerable to opioid-related complications due to multiple comorbidities, polypharmacy, and reduced clearance by the kidney of active drug metabolites. However, opioid use is common among patients receiving dialysis with estimates of use indicating that >60% receive an opioid prescription in a given year. In addition, over 20% of ESRD patients use opioids chronically, defined as >90 days in a calendar year. These rates of opioid prescription in the ESRD population are approximately three times that seen in the general Medicare population. Significant geographic variation in opioid prescriptions has been reported at both the state and dialysis facility [Bailie 2004] level.

In 2016, the CDC released guidelines for opioid prescription in an effort to ensure safe and effective treatment of chronic pain, while reducing the risk of addiction, overdose and death. These guidelines call for increased discussion and follow up between patients and providers, use of the lowest dose/duration possible, and consideration for non-opioid treatment modalities. Other recommendations note that depression, anxiety, and sleep disorders are associated with pain and should be considered in patient assessment.

Higher doses of opioids in the ESRD population have been associated with increased risk of falls and fractures compared to lower doses (which still impose some incremental risk) [Ishida, 2018]. Other authors, using USRDS data through 2010 reported that higher opioid doses correlated with death in a monotonically increasing fashion [Kimmel, 2017].

Co-prescription of benzodiazepines has been reported in 30% of opioid prescriptions [Ruchi 2019] in the ESRD population and increased the odds of hospitalization by 50%. The prevalence of opioid and benzodiazepine use in dialysis patients is highly variable between centers [Paramanandam 2011]. These findings suggest an opportunity exists for greater use of state Prescription Drug Monitoring Programs (PDMP), which have been demonstrated to reduce opioid MME doses [Change 2016] as well as opioid related mortality [Patrick, 2016].

Dialysis patients with chronic opioid prescriptions (>90 days) had increased mortality, dialysis discontinuation, and hospitalization when compared with patients without an opioid prescription [Kimmel, 2017]. However, when patients in the general population who receive chronic opioids also received a naloxone prescription, there were 47% fewer opioid-related ED visits per month in the 6 months after receipt of the naloxone prescription [Coffin, 2016].

In summary, there is evidence in the literature to link unsafe opioid prescription practices to serious adverse event, such as hospitalization and mortality, in the dialysis population. Furthermore, interventions such as use of PDMPs and co-prescription of naloxone have been demonstrated to reduce these risks.

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

The following search was conducted in PubMed in February 2019

("kidney failure, chronic"[MeSH Terms] OR ("kidney"[All Fields] AND "failure"[All Fields] AND "chronic"[All Fields]) OR "chronic kidney failure"[All Fields] OR "esrd"[All Fields]) AND ("analgesics, opioid"[Pharmacological Action] OR "analgesics, opioid"[MeSH Terms] OR ("analgesics"[All Fields] AND "opioid"[All Fields]) OR "opioid analgesics"[All Fields] OR "opioid"[All Fields])

This returned 268 articles that were reviewed and of these 43 were selected for presentation to the Technical Expert Panel that was convened to make recommendations regarding this measure. Articles relevant to the summary above are included in 3.1.6.3.

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

Daubresse M, Alexander GC, Crews DC, Segev DL, McAdams-DeMarco MA. Trends in Opioid Prescribing Among Hemodialysis Patients, 2007-2014. Am J Nephrol. 2019;49(1):20-31. doi: 10.1159/000495353. Epub 2018 Dec 13. PMID: 30544114; PMCID: PMC6341485.

Abstract

Background: Hemodialysis (HD) patients frequently experience pain. Previous studies of HD patients suggest increased opioid prescribing through 2010. It remains unclear if this trend continued after 2010 or declined with national trends.

Methods: Longitudinal cohort study of 484,745 HD patients in the United States Renal Data System/Medicare data. We used Poisson/negative binomial regression to estimate annual incidence rates of opioid prescribing between 2007 and 2014. We compared prescribing rates with the general US population using IQVIA's National Prescription Audit data. Outcomes included the following: percent of HD patients receiving an opioid prescription, rate of opioid prescriptions, quantity, days' supply, morphine milligram equivalents (MME) dispensed per 100 person-days, and prescriptions per person.

Results: In 2007, 62.4% of HD patients received an opioid prescription. This increased to 63.2% in 2010 then declined to 53.7% by 2014. Opioid quantity peaked in 2011 at 73.5 pills per 100 person-days and declined to 62.6 pills per 100 person-days in 2014. MME peaked between 2010 and 2012 then declined through 2014. In 2014, MME rates were 1.8-fold higher among non-Hispanic patients and 1.6-fold higher among low-income patients. HD patients received 3.2-fold more opioid prescriptions per person compared to the general US population and were primarily prescribed oxycodone and hydrocodone. Between 2012 and 2014, HD patients experienced greater declines in opioid prescriptions per person (18.2%) compared to the general US population (7.1%).

Conclusion: Opioid prescribing among HD patients declined between 2012 and 2014. However, HD patients continue receiving substantially more opioids than the general US population.

Ruchi R, Bozorgmehri S, Ozrazgat-Baslanti T, Segal MS, Shukla AM, Mohandas R, Kumar S. Opioid Safety and Concomitant Benzodiazepine Use in End-Stage Renal Disease Patients. Pain Res Manag. 2019 Oct 20;2019:3865924. doi: 10.1155/2019/3865924. PMID: 31772694; PMCID: PMC6854236.

Abstract

Background. Opioid use is common in end-stage renal disease (ESRD) patients. However, safety of individual opioids and concomitant benzodiazepine use has not been studied. Objective. To study the epidemiology of opioid and concomitant benzodiazepine use in ESRD population. To study the clinical safety profile of individual opioids in patients on hemodialysis. Design. Retrospective analysis of the U.S. Renal Data System. A comprehensive review of the current literature was performed to update currently used opioid safety classification. Participants. ESRD patients ≥18 years on hemodialysis who were enrolled in Medicare A and B

and Part D between 2006 and 2012, excluding those with malignancy. Main Measures. Hospital admission with diagnosis of prescription opioid overdose within 30, 60, and 90 days of prescription; death due to opioid overdose. Results. Annually, the percentage of patients prescribed any opioid was 52.2%. Overall trend has been increasing except for a small dip in 2011, despite which the admissions due to opioid overdose have been rising. 30% of those who got a prescription for opioids also got a benzodiazepine prescription. 56.5% of these patients received both prescriptions within a week of each other. Benzodiazepine use increased the odds of being on opioids by 3.27 (CI 3.21–3.32) and increased the odds of hospitalization by 50%. Opioids considered safe such as fentanyl and methadone were associated with 3 and 6 folds higher odds of hospitalization within 30 days of prescription. Hydrocodone had the lowest odds ratio (1.9, CI 1.8–2.0). Conclusions. Concurrent benzodiazepine use is common and associated with higher risk of hospitalization due to opioid overdose. Possible opioid-associated hospital admission rate is 4-5 times bigger in ESRD population than general population. Current safety classification of opioids in these patients is misleading, and even drugs considered safe based on pharmacokinetic data are associated with moderate to very high risk of hospitalization. We propose a risk-stratified classification of opioids and suggest starting to use them in all ESRD patients.

Patrick SW, Fry CE, Jones TF, Buntin MB. Implementation Of Prescription Drug Monitoring Programs Associated With Reductions In Opioid-Related Death Rates. Health Aff (Millwood). 2016 Jul 1;35(7):1324-32. doi: 10.1377/hlthaff.2015.1496. Epub 2016 Jun 22. PMID: 27335101; PMCID: PMC5155336.

Abstract

Over the past two decades the number of opioid pain relievers sold in the United States rose dramatically. This rise in sales was accompanied by an increase in opioid-related overdose deaths. In response, forty-nine states (all but Missouri) created prescription drug monitoring programs to detect high-risk prescribing and patient behaviors. Our objectives were to determine whether the implementation or particular characteristics of the programs were effective in reducing opioid-related overdose deaths. In adjusted analyses we found that a state's implementation of a program was associated with an average reduction of 1.12 opioid-related overdose deaths per 100,000 population in the year after implementation. Additionally, states whose programs had robust characteristics-including monitoring greater numbers of drugs with abuse potential and updating their data at least weekly-had greater reductions in deaths, compared to states whose programs did not have these characteristics. We estimate that if Missouri adopted a prescription drug monitoring program and other states enhanced their programs with robust features, there would be more than 600 fewer overdose deaths nationwide in 2016, preventing approximately two deaths each day.

Chang HY, Lyapustina T, Rutkow L, Daubresse M, Richey M, Faul M, Stuart EA, Alexander GC. Impact of prescription drug monitoring programs and pill mill laws on high-risk opioid prescribers: A comparative interrupted time series analysis. Drug Alcohol Depend. 2016 Aug 1;165:1-8. doi: 10.1016/j.drugalcdep.2016.04.033. Epub 2016 Jun 2. PMID: 27264166; PMCID: PMC4985620.

Abstract

Background: Prescription drug monitoring programs (PDMPs) and pill mill laws were implemented to reduce opioid-related injuries/deaths. We evaluated their effects on high-risk prescribers in Florida.

Methods: We used IMS Health's LRx Lifelink database between July 2010 and September 2012 to identify opioid-prescribing prescribers in Florida (intervention state, N: 38,465) and Georgia (control state, N: 18,566). The pre-intervention, intervention, and post-intervention periods were: July 2010-June 2011, July 2011-September 2011, and October 2011-September 2012. High-risk prescribers were those in the top 5th percentile of opioid volume during four consecutive calendar quarters. We applied comparative interrupted

time series models to evaluate policy effects on clinical practices and monthly prescribing measures for low-risk/high-risk prescribers.

Results: We identified 1526 (4.0%) high-risk prescribers in Florida, accounting for 67% of total opioid volume and 40% of total opioid prescriptions. Relative to their lower-risk counterparts, they wrote sixteen times more monthly opioid prescriptions (79 vs. 5, p<0.01), and had more prescription-filling patients receiving opioids (47% vs. 19%, p<0.01). Following policy implementation, Florida's high-risk providers experienced large relative reductions in opioid patients and opioid prescriptions (-536 patients/month, 95% confidence intervals [CI] -829 to -243; -847 prescriptions/month, CI -1498 to -197), morphine equivalent dose (-0.88mg/month, CI -1.13 to -0.62), and total opioid volume (-3.88kg/month, CI -5.14 to -2.62). Low-risk providers did not experience statistically significantly relative reductions, nor did policy implementation affect the status of being high- vs. low- risk prescribers.

Conclusions: High-risk prescribers are disproportionately responsive to state policies. However, opioids-prescribing remains highly concentrated among high-risk providers.

Coffin PO, Behar E, Rowe C, Santos GM, Coffa D, Bald M, Vittinghoff E. Nonrandomized Intervention Study of Naloxone Coprescription for Primary Care Patients Receiving Long-Term Opioid Therapy for Pain. Ann Intern Med. 2016 Aug 16;165(4):245-52. doi: 10.7326/M15-2771. Epub 2016 Jun 28. PMID: 27366987; PMCID: PMC5783639.

Abstract

Background: Unintentional overdose involving opioid analgesics is a leading cause of injury-related death in the United States.

Objective: To evaluate the feasibility and effect of implementing naloxone prescription to patients prescribed opioids for chronic pain.

Design: 2-year nonrandomized intervention study.

Setting: 6 safety-net primary care clinics in San Francisco, California.

Participants: 1985 adults receiving long-term opioid therapy for pain.

Intervention: Providers and clinic staff were trained and supported in naloxone prescribing.

Measurements: Outcomes were proportion of patients prescribed naloxone, opioid-related emergency department (ED) visits, and prescribed opioid dose based on chart review.

Results: 38.2% of 1985 patients receiving long-term opioids were prescribed naloxone. Patients prescribed higher doses of opioids and with an opioid-related ED visit in the past 12 months were independently more likely to be prescribed naloxone. Patients who received a naloxone prescription had 47% fewer opioid-related ED visits per month in the 6 months after receipt of the prescription (incidence rate ratio [IRR], 0.53 [95% CI, 0.34 to 0.83]; P = 0.005) and 63% fewer visits after 1 year (IRR, 0.37 [CI, 0.22 to 0.64]; P < 0.001) compared with patients who did not receive naloxone. There was no net change over time in opioid dose among those who received naloxone and those who did not (IRR, 1.03 [CI, 0.91 to 1.27]; P = 0.61).

Limitation: Results are observational and may not be generalizable beyond safety-net settings.

Conclusion: Naloxone can be coprescribed to primary care patients prescribed opioids for pain. When advised to offer naloxone to all patients receiving opioids, providers may prioritize those with established risk factors. Providing naloxone in primary care settings may have ancillary benefits, such as reducing opioid-related adverse events.

Ishida JH, McCulloch CE, Steinman MA, Grimes BA, Johansen KL. Opioid Analgesics and Adverse Outcomes among Hemodialysis Patients. Clin J Am Soc Nephrol. 2018 May 7; 13(5):746-753. doi: 10.2215/CJN.09910917 Epub 2018 Apr 19.

ABSTRACT

BACKGROUND AND OBJECTIVES:

Patients on hemodialysis frequently experience pain and may be particularly vulnerable to opioid-related complications. However, data evaluating the risks of opioid use in patients on hemodialysis are limited.

DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS:

Using the US Renal Data System, we conducted a cohort study evaluating the association between opioid use (modeled as a time-varying exposure and expressed in standardized oral morphine equivalents) and time to first emergency room visit or hospitalization for altered mental status, fall, and fracture among 140,899 Medicare-covered adults receiving hemodialysis in 2011. We evaluated risk according to average daily total opioid dose (>60 mg, ≤60 mg, and per 60-mg dose increment) and specific agents (per 60-mg dose increment).

RESULTS:

The median age was 61 years old, 52% were men, and 50% were white. Sixty-four percent received opioids, and 17% had an episode of altered mental status (15,658 events), fall (7646 events), or fracture (4151 events) in 2011. Opioid use was associated with risk for all outcomes in a dose-dependent manner: altered mental status (lower dose: hazard ratio, 1.28; 95% confidence interval, 1.23 to 1.34; higher dose: hazard ratio, 1.67; 95% confidence interval, 1.56 to 1.78; hazard ratio, 1.29 per 60 mg; 95% confidence interval, 1.26 to 1.33), fall (lower dose: hazard ratio, 1.28; 95% confidence interval, 1.21 to 1.36; higher dose: hazard ratio, 1.45; 95% confidence interval, 1.31 to 1.61; hazard ratio, 1.04 per 60 mg; 95% confidence interval, 1.03 to 1.05), and fracture (lower dose: hazard ratio, 1.44; 95% confidence interval, 1.33 to 1.56; higher dose: hazard ratio, 1.65; 95% confidence interval, 1.44 to 1.89; hazard ratio, 1.04 per 60 mg; 95% confidence interval, 1.04 to 1.05). All agents were associated with a significantly higher hazard of altered mental status, and several agents were associated with a significantly higher hazard of fall and fracture.

CONCLUSIONS:

Opioids were associated with adverse outcomes in patients on hemodialysis, and this risk was present even at lower dosing and for agents that guidelines have recommended for use.

Kimmel PL, Fwu CW, Abbott KC, Eggers AW, Kline PP, Eggers PW. Opioid Prescription, Morbidity, and Mortality in United States Dialysis Patients. J Am Soc Nephrol. 2017 Dec;28(12):3658-3670. doi: 10.1681/ASN.2017010098. Epub 2017 Sep 21.

ABSTRACT

Aggressive pain treatment was advocated for ESRD patients, but new Centers for Disease Control and Prevention guidelines recommend cautious opioid prescription. Little is known regarding outcomes associated with ESRD opioid prescription. We assessed opioid prescriptions and associations between opioid prescription and dose and patient outcomes using 2006-2010 US Renal Data System information in patients on maintenance dialysis with Medicare Part A, B, and D coverage in each study year (*n*=671,281, of whom 271,285 were unique patients). Opioid prescription was confirmed from Part D prescription claims. In the 2010 prevalent cohort (*n*=153,758), we examined associations of

opioid prescription with subsequent all-cause death, dialysis discontinuation, and hospitalization controlled for demographics, comorbidity, modality, and residence. Overall, >60% of dialysis patients had at least one opioid prescription every year. Approximately 20% of patients had a chronic (≥90-day supply) opioid prescription each year, in 2010 usually for hydrocodone, oxycodone, or tramadol. In the 2010 cohort, compared with patients without an opioid prescription, patients with short-term (1-89 days) and chronic opioid prescriptions had increased mortality, dialysis discontinuation, and hospitalization. All opioid drugs associated with mortality; most associated with worsened morbidity. Higher opioid doses correlated with death in a monotonically increasing fashion. We conclude that opioid drug prescription is associated with increased risk of death, dialysis discontinuation, and hospitalization in dialysis patients. Causal relationships cannot be inferred, and opioid prescription may be an illness marker. Efforts to treat pain effectively in patients on dialysis yet decrease opioid prescriptions and dose deserve consideration.

Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain--United States, 2016. JAMA. 2016 Apr 19;315(15):1624-45. doi: 10.1001/jama.2016.1464. Review.

ABSTRACT

This guideline provides recommendations for primary care clinicians who are prescribing opioids for chronic pain outside of active cancer treatment, palliative care, and end-of-life care. The guideline addresses 1) when to initiate or continue opioids for chronic pain; 2) opioid selection, dosage, duration, follow-up, and discontinuation; and 3) assessing risk and addressing harms of opioid use. CDC developed the guideline using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework, and recommendations are made on the basis of a systematic review of the scientific evidence while considering benefits and harms, values and preferences, and resource allocation. CDC obtained input from experts, stakeholders, the public, peer reviewers, and a federally chartered advisory committee. It is important that patients receive appropriate pain treatment with careful consideration of the benefits and risks of treatment options. This guideline is intended to improve communication between clinicians and patients about the risks and benefits of opioid therapy for chronic pain, improve the safety and effectiveness of pain treatment, and reduce the risks associated with long-term opioid therapy, including opioid use disorder, overdose, and death. CDC has provided a checklist for prescribing opioids for chronic pain (http://stacks.cdc.gov/view/cdc/38025) as well as a website (http://www.cdc.gov/drugoverdose/prescribingresources.html) with additional tools to guide clinicians in implementing the recommendations.

Lentine KL, Yuan H, Tuttle-Newhall JE, Xiao H, Chawa V, Axelrod D, Brennan DC, Dharnidharka VR, Beuer C, Schnitzler MA. Quantifying prognostic impact of prescription opioid use before kidney transplantation through linked registry and pharmaceutical claims data. Transplantation. 2015 Jan;99(1):187-96. doi: 10.1097/TP.0000000000000248.

ABSTRACT

BACKGROUND: Limited data are available on the outcome implications of prescription narcotic use before kidney transplantation.

METHODS: We examined a novel database wherein national transplant registry identifiers for kidney transplant recipients were linked to records from a large U.S. pharmaceutical claims clearinghouse (2005-2010). We selected recipients with 1 year of captured pretransplant pharmaceutical fill records (N=31,197). Opioid analgesic fills in the year before transplantation were normalized to morphine

equivalents (ME) and expressed as mg/kg exposures. Adjusted associations of ME level with posttransplant graft and patient survival (adjusted hazards ratio, aHR) were quantified by multivariate Cox regression.

RESULTS: Among the 29% of the sample who filled opioid prescriptions in the year before transplantation, the 25th, 50th, and 75th percentiles of annual ME were 1.8, 5.5, and 23.7 mg/kg, respectively. Three-year graft survival was 88.0% and 84.4% in live donor recipients with upper quartiles of ME use, compared with 92.0% among those who did not receive prescription narcotics (P<0.0001). Adjusted risks of posttransplant death and all-cause graft loss in live donor recipients with the highest quartile of narcotic use were 2.3 times (aHR, 2.27; 95% confidence interval, 1.66-3.10) and 1.8 times (aHR, 1.75; 95% confidence interval, 1.37-2.26), respectively, that of narcotic nonusers. Graded associations of pretransplant opioid exposure level with death and graft loss after deceased donor transplantation were also observed.

CONCLUSIONS: Although associations may in part reflect underlying conditions or behaviors, high levels of prescription opioid use before kidney transplantation predict increased risk of posttransplant death and graft loss.

Willy ME, Graham DJ, Racoosin JA, Gill R, Kropp GF, Young J, Yang J, Choi J, MaCurdy TE, Worrall C, Kelman JA. Candidate metrics for evaluating the impact of prescriber education on the safe use of extended-release/long-acting (ER/LA) opioid analgesics. Pain Med. 2014 Sep;15(9):1558-68. doi: 10.1111/pme.12459. Epub 2014 May 15.

ABSTRACT

OBJECTIVE: The objective of this study was to develop metrics to assess opioid prescribing behavior as part of the evaluation of the Extended-Release/Long-Acting (ER/LA) Opioid Analgesic Risk Evaluation and Mitigation Strategies (REMS).

DESIGN: Candidate metrics were selected using published guidelines, examined using sensitivity analyses, and applied to cross-sectional rolling cohorts of Medicare patients prescribed with extended-release oxycodone (ERO) between July 2, 2006 and July 1, 2011. Potential metrics included prescribing opioid-tolerant-only ER/LA opioid analgesics to non-opioid-tolerant patients, prescribing early fills to patients, and ordering drug screens.

RESULTS: Proposed definitions for opioid tolerance were seven continuous days of opioid usage of at least 30 mg oxycodone equivalents, within the 7 days (primary) or 30 days (secondary) prior to first opioid-tolerant-only ERO prescription. Forty-four percent of opioid-tolerant-only ERO episodes met the primary opioid tolerance definition; 56% met the secondary definition. Fills were deemed "early" if a prescription was filled before 70% (primary) or 50% (secondary) of the prior prescription's days' supply was to be consumed. Five percent (primary) and 2% (secondary) of episodes had more than or equal to two early fills during treatment. At least one drug screen was billed in 14% of episodes. Stratified analyses indicated that older patients were less likely to be opioid tolerant at the time of the first opioid-tolerant-only ERO prescription.

CONCLUSIONS: Investigators propose three metrics to monitor changes in prescribing behaviors for opioid analgesics that might be used to evaluate the ER/LA Opioid Analgesics REMS. Low frequencies of patients, particularly those >85 years, were likely to be opioid tolerant prior to receiving prescriptions for opioid-tolerant-only ERO.

Paramanandam G, Prommer E, Schwenke DC. Opioid and benzodiazepine use in end-stage renal disease: a systematic review. J Palliat Med. 2011 Sep;14(9):1029-33. doi: 10.1089/jpm.2011.0103. Epub 2011 Aug 8.

ABSTRACT

BACKGROUND AND OBJECTIVES: Chronic pain and psychiatric disorders are common in dialysis patients, but the extent to which opioids and benzodiazepines are used is unclear. We conducted a systematic review to determine the: (1) prevalence of opioid and benzodiazepine use among dialysis patients; (2) reasons for use; (3) effectiveness of symptom control; and (4) incidence of adverse events.

DESIGN, SETTING, PARTICIPANTS, & MEASUREMENTS: Two authors reviewed all relevant citations in MEDLINE/EMBASE/CINAHL/BIOSIS Previews/Cochrane and hand-searched bibliographies. Studies after 1990 reporting prevalence estimates for opioid and/or benzodiazepine use in ≥50 dialysis patients were included.

RESULTS: We identified 15 studies from 12 countries over 1995 to 2006. Sample size ranged from 75 to 12,782. Prevalence of opioid and benzodiazepine use was variable, ranging from 5 to 36% (95% CI, 4.1 to 45.5%; n=10) and 8 to 26% (95% CI, 7.1 to 27.3%; n=9), respectively. Prevalence was positively correlated with years on dialysis. Five studies reported on the same cohorts but gave different prevalence estimates. One study verified medication use through patient interviews. Reasons for use were reported in one study. Effectiveness of pain control varied from 17 to 38%, and 72 to 84% of patients with significant pain had no analgesia (n=2). No study rigorously examined for adverse events.

CONCLUSIONS: The prevalence of opioid and benzodiazepine use in dialysis patients is highly variable between centers. Further information is needed regarding the appropriateness of these prescriptions, adequacy of symptom control, and incidence of adverse effects in this population.

Bailie GR, Mason NA, Bragg-Gresham JL, Gillespie BW, Young EW. Analgesic prescription patterns among hemodialysis patients in the DOPPS: potential for under prescription. Kidney Int. 2004 Jun;65(6):2419-25.

ABSTRACT

BACKGROUND: Dialysis patients require special consideration regarding analgesics, given their altered pharmacokinetic and pharmacodynamic profiles and increased potential for adverse reactions.

METHODS: Analgesic prescription patterns were investigated using data from the Dialysis Outcomes and Practice Patterns Study (DOPPS), with 3749 patients in 142 United States facilities studied between May 1996 and September 2001.

RESULTS: The proportion of patients prescribed any analgesic decreased from 30.2% to 24.3%; narcotic prescriptions decreased from 18.0% to 14.9%. The most commonly prescribed narcotics were propoxyphene/acetaminophen combinations (47.2%). Combinations containing acetaminophen were prescribed concurrently for 84.1% of patients on narcotics. About one half of prescriptions for narcotics, acetaminophen, and cyclooxygenase-2 (COX-2) agents were for 12 months or more; one half of prescriptions for nonsteroidal anti-inflammatory drugs (NSAIDs) were for 8 months or more. The proportion of patients prescribed analgesics varied by facility (mean +/- SD = 27.9%+/- 18.9% for all analgesics, range 0% to 89.3%). Analgesic prescription was more likely among the elderly, women, and patients with cardiovascular disease (other than coronary artery disease or congestive heart failure),

lung and psychiatric disease, cancer (other than skin), and recurrent cellulitis. Patients prescribed laxatives were almost twice as likely to be on a narcotic (odds ratio = 1.95, P < 0.0001). Analgesic prescription did not correlate with loss of residual renal function or hospitalization for a gastrointestinal disorder. Three-quarters of patients reporting moderate to very severe pain were not prescribed analgesics. Furthermore, 74% of patients with pain that interfered with work had no analgesic prescription.

CONCLUSION: Dialysis patients and providers may benefit from both refinement of existing guidelines and a renewed understanding regarding appropriate prescription of analgesics.

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.*

Several observational studies have demonstrated an association between unsafe opioid use in the dialysis population and higher risk of fall/fracture, hospitalization, and mortality. Unsafe opioid use is typically defined as >50 morphine milligram equivalents (MME), duration > 90 days, or co-prescription with a benzodiazepine.

The measure focus is the process determining the percentage of all dialysis patients attributable to an opioid prescriber's group practice who had an unsafe opioid prescription written within the year.

The measure is risk adjusted to mitigate against the unintended consequences of under treatment of pain in patients with comorbidities that have a significant pain component (e.g., cancer, sickle cell disease). By adjusting for case-mix at the prescriber's group practice, our intent is for providers to be able to write necessary opioid prescriptions for patients with greater comorbidity, and likely greater analgesia needs, since the measure does not penalize individual prescribing events or hold providers to an absolute scale or threshold. Rather, the measure identifies the small number of group practices that, based on their year-long prescribing patterns, have extreme deviations relative to their peers.

Once implemented practitioner performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in reducing unsafe opioid use, while accounting for patients where higher dose or longer term therapy may be warranted.

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

Analysis of January 2017 – December 2017 data indicate the physician level mean percentage of patient months with unsafe opioid use is 39.7% (Std Dev 19.8%). Distribution of performance scores:

Min=0, Max=100, Median=38.5, Interquartile range= [25, 52.6].

N of prescriber groups=5,123, N of patients= 204,034.

Of the ESRD patients who are prescribed an opioid, 39.7% of those prescriptions met the above definition for unsafe use. Given that the interquartile range extends from 25% up to 52.6.2%, there is significant variation in provider group performance indicating that a performance gap exists that may be modifiable.

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.

N/A

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.*

Using data from January-December 2017: age, sex, race, ethnicity, dialysis vintage, employment status, Medicare coverage, and Area Deprivation Index (ADI) were evaluated in a logistic regression model for unsafe opioid use. Data on patient level SDS/SES factors were obtained from Medicare claims and administrative data; zip code level data for the Area Deprivation Index (ADI) are obtained from Census data (2009-2013), based on patient zip-code. Below are the odds ratios for these patient characteristics.

Age, sex, race, and ethnicity are all statistically significant predictors of unsafe opioid use. Specifically, patients under age 25 had 17% higher odds of having unsafe opioid use for each year increase in age. Between ages 25 and 65 the odds decreased with age (0.7% each year) with a sharper decline after age 65 (2.7% each year). Females had a 5% higher odds of having unsafe opioid use versus males. Hispanic ethnicity was associated with lower odds of Opioid unsafe use whereas Black race had a 35% lower odds of unsafe opioid use compared to whites. Unemployment or "other" employment status as well as dual eligible status were all associated with higher odds of Opioid unsafe use. The analysis results for age, race, sex and patient SES indicate potential disparities in unsafe opioid use. Patient-level SDS/SES variables are not included as adjustments in the measure since, in the absence of biological effects explaining these differences, risk adjustment for these factors could potentially mask disparities in care. However, these variables do highlight certain subgroups that may be at higher risk for unsafe opioid use as prescribers consider interventions to close performance gaps.

Odds ratio of having unsafe opioid use:

Age:

For the continuous age, the Odds Ratio (95% CI) is 1.17 (1.07, 1.27), P-value is <0.001.

For the age spline at 25 years, the Odds Ratio (95% CI) is 0.85 (0.78, 0.93), P-value is <0.001.

For the age spline at 65 years, the Odds Ratio (95% CI) is 0.98 (0.98, 0.98), P-value is <0.001.

Sex:

For Female: the Odds Ratio (95% CI) is 1.05 (1.02, 1.07), P-value is <0.001.

Male was used as the reference group.

Race:

White was used as the reference group.

For Black: the Odds Ratio (95% CI) is 0.68 (0.66, 0.7), P-value is <0.001.

For Other race: the Odds Ratio (95% CI) is 0.54 (0.52, 0.58), P-value is <0.001.

Ethnicity:

For Hispanic: the Odds Ratio (95% CI) is 0.65 (0.62, 0.68), P-value is <0.001.

Non-Hispanic was used as the reference group.

Employment Status:

Employed was used as the reference group.

For Unemployed: The Odds Ratio (95% CI) is 1.15 (1.11, 1.19), and the P-value is <0.001.

For Other: The Odds Ratio (95% CI) is 1.23 (1.19, 1.27), and the P-value is <0.001.

Medicare Coverage:

Dual eligibility: the Odds Ratio (95% CI) is 1.15 (1.13, 1.18), and the P-value is <0.001.

Non-Dual eligibility was used as the reference group.

ADI (zipcode-level):

National percentile ADI score: The Odds Ratio (95% CI) is 1.00 (1.00, 1.00), and the P-value is 0.047.

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

N/A

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.)*

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: Prescriber_Group_MCP_Group_Data_Dictionary_Code_List-637454373896340050.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.

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

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

The numerator is the number of patients in the denominator who were prescribed an opioid that was either >90 days duration during the year, >50 MME, or overlapped in time with a benzodiazepine prescription.

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 days duration for an opioid prescription is determined by adding up the days supplied from the Medicare Part D claim file over the entire one year reporting period. When there are overlapping days of opioid use, the number of distinct days of an opioid prescription are counted and in the case of multiple prescribers, each opioid is only attributable to one prescriber group. The number of MME is calculated based on conversion factors for each opioid medication. Patients will be counted if there is a prescription for >50 MME, regardless of the duration of that prescription. Dates for benzodiazepines prescriptions are also obtained from Medicare Part D claims and then compared to the dates of opioid coverage. If there is any overlap of dates, regardless of prescriber, the patient will be counted in the numerator.

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

The denominator is the number of patients associated with an opioid prescriber's group practice who are receiving maintenance dialysis (in-center or home dialysis) for any duration who receive an opioid prescription during the one year reporting 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).*

For each dialysis patient, we identify the opioid prescriber from Part D Medicare claims. Medicare Provider files are used to determine a prescriber's group partners by linking the individual National Provider Identifier (NPI) numbers with a shared active Tax Identification Number (TIN) for the group. Medicare dialysis claims are used to identify patients that are receiving in-center or home dialysis during the reporting period.

If a patient receives an opioid prescription from more than one provider group during the reporting period, the patient will be counted in each group for the corresponding opioid prescriptions.

Opioid prescription use is censored at the time of kidney transplantation, discontinuation of dialysis, or the end of the reporting period.

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

Patients who have a hospice claim at any time (either before or after the opioid prescription date) during the one year reporting period are excluded.

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 status is determined from final action claims submitted to CMS by Hospice providers. Hospice related claims include beneficiaries in both Medicare fee-for-service and Medicare managed care plans. Patients are identified as receiving hospice care if they have any final action claims submitted to Medicare by hospice providers in the current month.

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.)*

N/A

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

Statistical risk model

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.)*

Flowchart provided in Appendix

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

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

N/A

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

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

N/A

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

If other, please describe in S.18.

Claims, Other, Registry 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.

CROWNWeb, Medicare Claims and the CMS Medical Evidence form 2728 are used as the data sources for establishing the denominator. Medicare Part D Claims are used for both the numerator and denominator. Medicare claims during the reporting period are used for the hospice exclusion criteria. Medicare claims from the year prior to the reporting period are used for comorbidity condition adjustments. The Medicare Provider Files from the CMS Integrated Data Repository (IDR) are used to identify practitioner's group partners.

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

No data collection instrument provided

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

Clinician: Group/Practice

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

Other

If other: Dialysis Facility

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

N/A

2. Validity – See attached Measure Testing Submission Form

Prescriber_Group_Testing_01052021-637469270642789642.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*)**:**
Measure Title: Unsafe Opioid Prescriptions at the Prescriber Group Level
Date of Submission: **1/5/2021**
Type of Measure:

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

*cell intentionally left blank

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

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

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

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

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*).

CROWNWeb, Medicare Claims and the CMS Medical Evidence form 2728 are used as the data sources for establishing the denominator. Medicare Part D Claims are used for both the numerator and denominator. Medicare claims are used for the hospice exclusion criteria and comorbidity condition adjustments. The Medicare Provider Files from the CMS Integrated Data Repository (IDR) are used to identify practitioner's group partners.

1.3. What are the dates of the data used in testing? January-December 2017

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

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

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

Patients on either home or in-center hemodialysis during the last HD treatment of the month from January-December 2017 were included in the analyses. The number of patients within each provider group ranged from 11-2411, with an average of 40 patients per group.

Public reporting of this measure on DFC or in the ESRD QIP would be restricted to physician groups with at least 11 eligible patients throughout the year for the measure. We have applied this restriction to all the reliability and validity testing reported here.

There are totally 103,157 physicians associated with 5123 physician groups, ranging from 1 to 2328 physicians per group with an average of 20 physicians per group.

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

There were a total of 204,034 eligible patients. Among those patient-months over the whole year, the average age was 61.2 years, 50.0 % of patient-months were female, 53.7 % were white, 41.0 % were black, 5.3 % reported race as "other", 15.1 % were Hispanic and 47.6 % had type II diabetes as the primary cause of ESRD.

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.

N/A

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.

Patient level:

- Employment status 6 months prior to ESRD
- Race
- Sex

- Ethnicity
- Medicare coverage*

**Assessed at a specific time point (e.g., at the reporting month). Medicare coverage in model was defined as:*

Medicare as primary and Medicaid

Medicare as primary and NO Medicaid

Medicare as secondary or Medicare HMO (e.g., Medicare Advantage)

Non-Medicare/missing

Data on patient level SDS/SES factors obtained from Medicare claims and administrative data.

ZIP code level – Area Deprivation Index (ADI) elements from Census data:

- Unemployment rate (%)
- Median family income
- Income disparity
- Families below the poverty level (%)
- Single-parent households with children <18 years old (%)
- Home ownership rate (%)
- Median home value
- Median monthly mortgage
- Median gross rent
- Population (aged 25+) with <9 years of education (%)
- Population (aged 25+) without high school diploma (%)

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*)

We used January-December 2017 Medicare Part D claims to calculate prescriber group -level annual performance scores. Our approach for determining measure reliability aligns with one-way analysis of variance (ANOVA), in which the between- prescriber group variation (σ_b^2) and the within- prescriber group variation ($\sigma_{t,w}^2$) in the measure is determined. The inter-unit reliability (IUR) measures the proportion of the total variation of a measure ($\sigma_b^2 + \sigma_{t,w}^2$) that is attributable to the between- prescriber group variation, the true signal reflecting the differences across prescriber groups. We assessed reliability by calculating inter-unit reliability (IUR) for the annual performance scores. If the measure were an average of individuals' measurements under the care of one prescriber group, the usual ANOVA approach would be used. The yearly based measure, however, is not a simple average and we instead estimate the IUR using a bootstrap approach, which uses a resampling scheme to estimate the within prescriber group variation that cannot be directly estimated by ANOVA. A small IUR (near 0) reveals that most of the variation of the measures between prescriber groups is driven by random noise, indicating the measure would not be a good characterization of

the differences among prescriber groups, whereas a large IUR (near 1) indicates that most of the variation between prescriber groups is due to the real difference between prescriber groups.

Here we describe our approach to calculating IUR. Let T1,…,TN be the annual rate of unsafe opioid prescriptions for N prescriber groups. To generate re-sampled data, we randomly draw patients from the national population B times (we set B=100). Using each re-sampled dataset, for the ith prescriber group, we calculate an annual rate ($T^*_{i,1}$,…, $T^*_{i,B}$) and their sample variance S_i^{*2}. From this it can be seen that

$$s^2_{t,w} = \frac{\sum_{i=1}^{N}\left[(n_i - 1)S_i^{*2}\right]}{\sum_{i=1}^{N}(n_i - 1)}$$

is a bootstrap estimate of the within- prescriber group variance in the catheter rate, where n_i is the number of subjects in the ith prescriber group. Calling on formulas from the one-way ANOVA, the total variation in the annual rate (i.e., $\sigma_b^2 + \sigma_{t,w}^2$) can be estimated by

$$s_t^2 = \frac{1}{n'(N-1)}\sum_{i=1}^{N} n_i(T_i - \bar{T})^2$$

where the overall weighted average of rate is $\bar{T} = \Sigma\ ni\ Ti\ /\ \Sigma\ ni,,$ and

$$n' = \frac{1}{N-1}\left(\sum n_i - \sum n_i^2 \Big/ \sum n_i\right)$$

is approximately the average prescriber group size (number of patients per prescriber group). Thus, the IUR = $\sigma_b^2/(\sigma_b^2 + \sigma_{t,w}^2)$ can be estimated by $(s_t^2 - s_{t,w}^2)/s_t^2$.

The reliability calculation only included prescriber groups with at least 11 patients during the entire year.

One limitation with the IUR is that, when many provider groups have outcomes around the national norm, the IUR can be small, though the measure can identify groups with extreme outcomes. To complement the IUR and to further assess whether the measure can identify providers with extreme outcomes, we also computed the profile IUR (PIUR) [1-3]. The PIUR, based on the measure's ability to consistently flag extreme provider groups, was computed with a two-step approach: first, we evaluated the ability of a measure to consistently profile groups with extreme outcomes; second, we mapped this reflagging ability to an IUR value computed by assuming no outlier group providers. This value was defined to be the PIUR.

Specifically, we considered a sample-splitting approach: within each provider group, we randomly split patients into two equally sized subgroups. For a given threshold (e.g., p-value<0.05), we determined whether a provider group was identified as extreme based on the first and second subgroup of patients. We repeated this process 100 times to estimate the probability that, given a provider group was classified as extreme based on the first subgroup of patients, it was also classified as extreme based on the second patient subgroup. This empirical reflagging rate was calibrated so as to identify an IUR value that would have yielded the same reflagging rate if the data had been hypothetically assumed to have no outlier provider groups. The identified IUR value would be the PIUR. If there were indeed no outlier group providers, IUR and PIUR would be equal. However, the difference between them, e.g. when the PIUR was substantially larger than the IUR, would indicate the data might have many outlier or extreme group providers that were not captured by the IUR itself.

1. He K, Dahlerus C, Xia L, Li Y, Kalbfleisch JD. The profile inter-unit reliability. Biometrics. 2019 Oct 23. doi: 10.1111/biom.13167. [Epub ahead of print]

2. Kalbfleisch JD, He K, Xia L, Li Y. Does the inter-unit reliability (IUR) measure reliability?, Health Services and Outcomes Research Methodology, 2018 Sept. 18(3), 215-225. Doi: 10.1007/s10742-018-0185-4.

3. He K, Kalbfleisch JD, Yang Y, Fei Z. Inter-unit reliability for nonlinear models. Stat Med. 2019 Feb 28;38(5):844-854. doi: 10.1002/sim.8005. Epub 2018 Oct 18.

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*)

The overall IUR is 0.86, which means 86% of the total variation of this prescriber group level measure can be explained by the differences among prescribers and not by random noise. To assess further whether the measure can identify prescriber groups with extreme values, we computed the PIUR, which is 0.98. The discrepancy between the IUR (0.86) and PIUR (0.98) indicates the existence of outlier prescriber groups that can be identified by the measure.

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

The value obtained for the IUR is high, while the PIUR deviates from the IUR. The results demonstrates that the measure can detect differences in performance scores across provider groups as well as outlier groups.

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)*

Validity of the measure was tested by evaluating the concordance between the prescriber group level measure scores, hospitalization metrics, and mortality rate. The justification of our test is based on several observational studies which have demonstrated an association between unsafe opioid use in the dialysis population and higher risk of hospitalization, and mortality. Specifically, we hypothesize that the lowest tertile,

T1 which has the highest proportion of unsafe opioid use, will have higher rates of hospitalization and mortality. Based on the literature reviewed, we expect this to be a moderately strong association.

We first conduct the test for the hospitalization outcomes. We divide practitioner groups, based on their measure scores, into 3 tertile classes (T1 to T3), and within each tertile class, we compute the hospitalization rates and average number of total days in the hospital in 2017. We then apply the Cochran-Armitage trend test to test the concordance between the tertile grouping and these prescriber group-level outcomes.

We use a slightly different approach for testing its association with mortality. This is because the definition of chronic opioid use, which requires that patients survive at least 90 days, may introduce selection bias (e.g., those who survived may be healthier) if we directly compare the tertile grouping with the average mortality rate within each tertile group. In fact, by doing so, we observed a reverse trend between unsafe opioid use and mortality rate. A more reasonable statistical approach is to stratify patients based on the length of time at risk during the performance period (1 month – 12 months) and then assess the association between mortality and use of opioids in each stratum. This way, we may be able to eliminate the selection bias.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Cut-points for the tertiles of the performance scores were defined as follows:

T1 (worst performance): 46.2%-99.9%
T2: 30.1%-46.3%
T3 (best performance): 0-30.1%

The patient hospitalization rate at the practitioner group level is 1.49, 1.46 and 1.41 for T1, T2, and T3 respectively (trend test $p<0.001$), while the average number of hospital days per year and patient at the practitioner group level is 6.1, 5.1 and 4.1 respectively (trend test $p<0.001$).

The practitioner group level average mortality rate is 0.19, 0.20, and 0.18 per patient-year for T1, T2 and T3 groups, respectively. Directly comparing the tertile grouping with the average mortality rate may yield biased results, as we stated in 2b1.2. Instead, we stratify patients based on the length of time at risk during the performance period (1 month – 12 months) and then assess the association between mortality and use of opioids in each stratum.

The table below shows the percentages of patient deaths by safe and unsafe Opioid use, stratified by months at risk. The results clearly show that, within each stratum (that is, the same at-risk set) unsafe use is associated with higher mortality.

N of months at risk	Opioid safe use, %deaths	Opioid unsafe use, %deaths
1m=1-30d	6.4%	7.1%
2m=31-60d	7.9%	9.5%
3m=61-90d	8.2%	10.7%
4m=91-120d	7.7%	11.1%
5m=121-150d	7.1%	10.5%
6m=151-180d	6.2%	10.8%
7m=181-210d	5.5%	9.8%
8m=211-240d	4.8%	9.1%

N of months at risk	Opioid safe use, %deaths	Opioid unsafe use, %deaths
9m=241-270d	3.8%	8.3%
10m=271-300d	2.7%	6.2%
11m=301-330d	2.0%	3.8%
12m=331-365d	0.7%	0.9%

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?*)
As hypothesized, the unsafe use of opioids is associated with more hospitalizations, longer length of hospital stay, and higher mortality, when stratified by time at risk. Specifically, when we compare similar time at risk, unsafe opioid use is associated with a 10-44% relative increase in the risk of death based on the number of months at risk. Taken together these results provide validation support for the measure in that lower rates of unsafe opioid use were associated with better performance on key outcomes.

_____

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*)
The following exclusions are applied to the denominator:

- Patients who have a hospice claim at any time (either before or after the opioid prescription date) during the reporting period are excluded.

The prescriber group level mean percentage of patients with an unsafe opioid prescription with and without the above exclusions are calculated and compared.

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*)

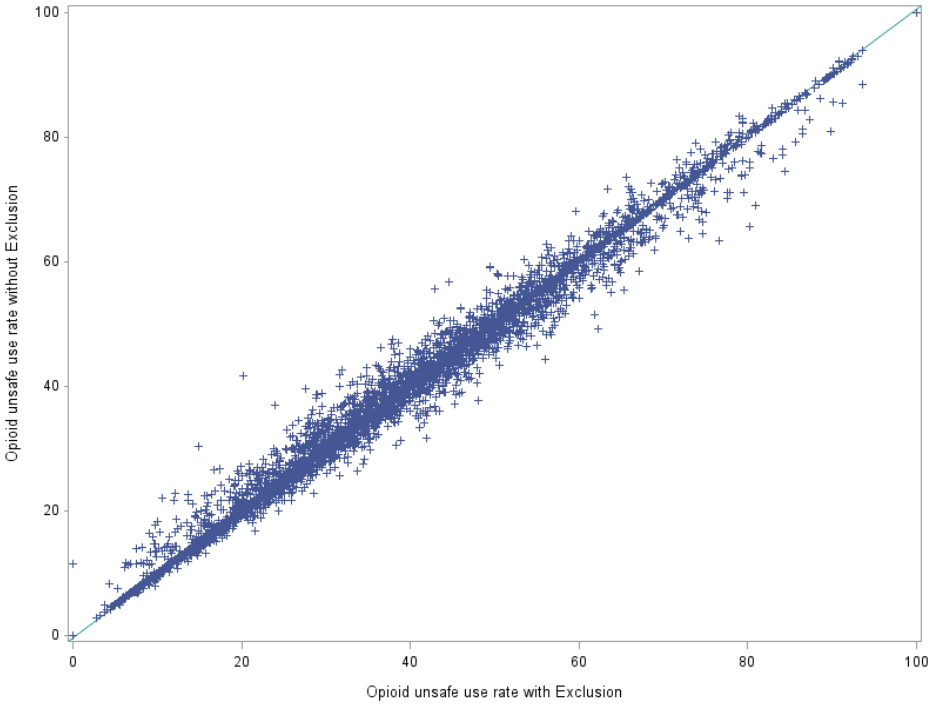
Table 1: Number and percent of unique patients excluded

Before Exclusion	After Exclusion	Percent
217,290	204,034	6.10%

Table 2: Distribution of performance scores before and after the exclusion

Opioid unsafe use rate	N	Mean	Standard Deviation	Minimum	Maximum
Before exclusion	5391	39.8	18.8	0	100
After exclusion	5123	39.3	18.8	0	100

Figure 1: Comparisons of Opioid measure values with and without exclusions



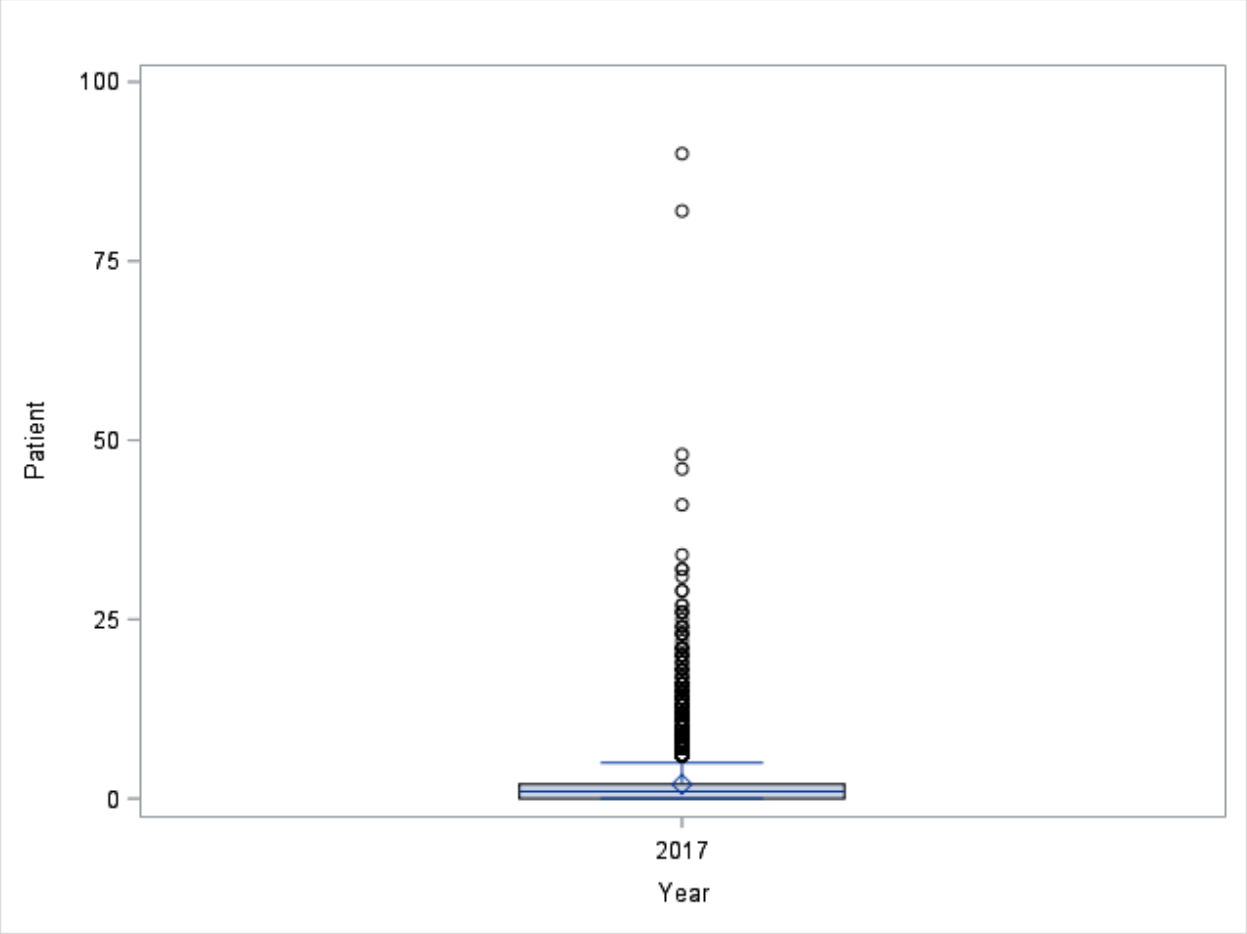
The correlation coefficient is 0.993 (p<.001).

Table 3. Comparison of performances with vs. without excluded patients

Opioid Unsafe use after Exclusion

Opioid Unsafe use before exclusion	excluded due to less than 11 eligible patients	Better than Expected	As Expected	Worse than Expected	Total
Better than Expected	8 (0.2%)	292 (5.4%)	37 (0.7%)	0	337 (6.3%)
As Expected	254 (4.7%)	18 (0.3%)	4587 ((85.1%)	16 (0.3%)	4875 (90.4%)
Worse than Expected	6 (0.1%)	0	10 (0.2%)	163 (3.0%)	179 (3.3%)
Total	268 (5.0%)	310 (5.8%)	4634 (86.0%)	179 (3.3%)	5391

Figure 2. Distribution of Excluded Patients at practitioner group level for 2017



2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (i.e., the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion)
The exclusion criteria are necessary since the percentage of patients excluded with each practitioner group is not evenly distributed across practitioners (Distribution shown in the boxplot). Due to the unequal distribution across practitioner groups, the exclusion criteria take into account that some practitioners treat a higher portion of patients with limited life expectancy. Additionally, our results shown in both the scatter-plot (Figure 1) as well as the Pearson Correlation Coefficient of 0.993 (p-value <0.0001) between the mean percentage of patients with Opioid unsafe use with and without the exclusion suggests that the overall impact of the exclusion on the measure's validity is not substantial since the two are highly correlated.

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

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

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 178 risk factors**

☐ **Stratification by risk categories**

☐ **Other**

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

The patient characteristics included in the model as covariates are:

Age: Age is included as a continuous variable, and two binary variables based on whether the patient is 25+ years old, or 65+ years old, respectively. Sex*

BMI at incidence

- BMI < 18.5
- 18.5 ≤ BMI < 25
- 25≤ BMI < 30
- BMI ≥30

- Duration of ESRD:
  - Less than one year
  - 1-2 years
  - 2-3 years
  - 3-6 years
  - 6+ years
- Nursing home status in previous year
  - None (0 days)
  - Short term (0-89 days)
  - Long term >=90 days)
- Diabetes as primary cause of ESRD
- Comorbidities at ESRD incidence:
  - Congestive heart failure
  - Atherosclerotic heart disease and other cardiac disease
  - Cerebrovascular disease, CVA, TIA
  - Peripheral vascular disease
  - Amputation
  - Diabetes other than as primary cause of ESRD (all types including diabetic retinopathy)
  - Chronic obstructive pulmonary disease
  - Inability to ambulate
  - Inability to transfer
  - Malignant neoplasm, cancer

- Tobacco use (current smoker)
- Alcohol dependence
- Drug dependence
- No Medical Evidence (CMS-2728) Form
- At least one of the comorbidities listed

A set of prevalent comorbidities based on Medicare inpatient claims (individual comorbidities categorized into 149 groups – see below)

*Denotes SDS/SES factor

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.
N/A

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?

In general, adjustment factors for this measure were selected based on several considerations. We began with selecting patient characteristics (listed above) that have been reported in the literature to be significant when considering opioid use in patients who are on dialysis and were supported by our Technical Expert Panel. Prior studies have indicated that younger patients, women, longer dialysis vintage, nursing home residence, and certain comorbidities are all associated with higher rates of opioid use in the dialysis population. These characteristics define our "base" model. Factors considered appropriate were then investigated with statistical models to determine if they were related to unsafe opioid use.

We then used the Agency for Healthcare Research and Quality (AHRQ) Clinical Classifications Software (CCS) diagnosis categories for prevalent comorbidity selection. First, we selected 241 of 283 prior year comorbidity groupers as potential candidates that had a prevalence greater than 0.1% in our population. Next, we used a stepwise variable selection approach (with a p-value cutoff of 0.01 in a logistic model) to identify 149 comorbidity variables that were associated with unsafe opioid use. More cutting edge machine learning techniques (such as LASSO) confirmed the results.

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?

Table 4a. Estimated Model Coefficients and p-values

Covariate	Estimate	Odds Ratio	P-value
Age	*	*	*
Continuous (years)	0.110	1.11605	0.00426
Spline at 14 years	-0.113	0.89358	0.00353
Spline at 60 years	-0.016	0.98418	0
Female	-0.036	0.96468	0.00473
BMI	*	*	*
<18.5	-0.014	0.98661	0.70942
18.5-24.9	reference	reference	reference
25-29.9	-0.003	0.99752	0.8819
>=30	0.074	1.07724	0
Time on ESRD	*	*	*
< 1 year	reference	reference	reference
1-2 years	0.332	1.39379	0
2-3 years	0.419	1.52083	0
3-6 years	0.426	1.53086	0
>=6 years	0.502	1.65176	0
Nursing home during the prior 365 days	*	*	*
No nursing home care (0 days)	reference	reference	reference
Short-term nursing home care (1-89 days)	-0.056	0.94583	0.0019
Long-term nursing home care (>=90 days)	0.210	1.23384	0
Cause of ESRD: Diabetes	-0.082	0.92127	0
Comorbidities at start of ESRD	*	*	*
Diabetes	-0.029	0.97137	0.1981
Congestive heart failure	-0.015	0.98501	0.34601
Coronary Artery Disease	0.038	1.0382	0.01822
Cerebrovascular disease, CVA, TIA	-0.061	0.94058	0.00978
Peripheral vascular disease	0.015	1.01505	0.49303
Amputation	0.072	1.0748	0.04277
Chronic obstructive pulmonary disease	0.144	1.15484	0
Tobacco use (current smoker)	0.080	1.08364	0.00065
Malignant neoplasm, Cancer	0.065	1.06686	0.0333
Alcohol dependence	0.056	1.05804	0.27343
Drug dependence	0.147	1.15791	0.00145
Inability to ambulate	0.096	1.10083	0.00806
Inability to transfer	-0.021	0.97936	0.69183
At least one of the comorbidities listed	0.030	1.03012	0.10142

Covariate	Estimate	Odds Ratio	P-value
No Medical Evidence (CMS-2728)	-0.062	0.9404	0.23439

*cell intentionally left blank

Table 4b. Prevalent Comorbidity Coefficients

Covariate	Estimate	Odds Ratio	P-value
Prevalent Comorbidities (condition groups)	*	*	*
Tuberculosis	-0.236	0.79017	0.00355
Septicemia (except in labor)	0.007	1.00728	0.69177
Mycoses	0.003	1.00269	0.91278
Hepatitis	0.074	1.07709	0.00087
Viral infection	-0.039	0.96167	0.18629
Sexually transmitted infections (not HIV or hepatitis)	-0.044	0.95728	0.55167
Immunizations and screening for infectious disease	-0.059	0.94307	0.00007
Cancer of other GI organs; peritoneum	-0.166	0.84717	0.20342
Cancer of bone and connective tissue	-0.056	0.94579	0.72901
Melanomas of skin	0.174	1.18948	0.06588
Other non-epithelial cancer of skin	0.036	1.03668	0.43383
Cancer of breast	-0.054	0.94749	0.20535
Cancer of cervix	-0.076	0.92688	0.32812
Cancer of ovary	0.203	1.22497	0.06527
Cancer of prostate	-0.096	0.90869	0.02665
Cancer of kidney and renal pelvis	0.061	1.06277	0.11126
Multiple myeloma	0.188	1.20688	0.00346
Cancer; other and unspecified primary	0.118	1.1255	0.04823
Secondary malignancies	0.267	1.30575	0.00001
Maintenance chemotherapy; radiotherapy	0.150	1.16145	0.03308
Benign neoplasm of uterus	-0.172	0.84182	0.0197
Other and unspecified benign neoplasm	-0.037	0.96348	0.1129
Thyroid disorders	0.067	1.06881	0.00002
Diabetes mellitus without complication	-0.058	0.94389	0.00018
Diabetes mellitus with complications	-0.022	0.97877	0.2216
Nutritional deficiencies	-0.028	0.9728	0.03977
Disorders of lipid metabolism	-0.069	0.93296	0
Gout and other crystal arthropathies	0.015	1.01469	0.43189
Other nutritional; endocrine; and metabolic disorders	-0.029	0.9714	0.03714
Deficiency and other anemia	-0.036	0.96504	0.28739
Sickle cell anemia	0.218	1.24329	0.00125
Coagulation and hemorrhagic disorders	-0.017	0.98303	0.30561
Diseases of white blood cells	-0.036	0.96438	0.10469

Covariate	Estimate	Odds Ratio	P-value
Other CNS infection and poliomyelitis	-0.150	0.86092	0.11721
Other hereditary and degenerative nervous system conditions	0.104	1.10958	0.00001
Paralysis	0.087	1.09055	0.02428
Headache; including migraine	-0.003	0.99752	0.90464
Coma; stupor; and brain damage	-0.004	0.99629	0.94144
Cataract	-0.044	0.95667	0.1218
Retinal detachments; defects; vascular occlusion; and retinopathy	0.001	1.00092	0.97866
Glaucoma	-0.014	0.98594	0.62747
Blindness and vision defects	0.008	1.00794	0.72267
Conditions associated with dizziness or vertigo	-0.057	0.94468	0.02205
Other ear and sense organ disorders	-0.091	0.91346	0.00084
Other nervous system disorders	0.252	1.28651	0
Heart valve disorders	-0.042	0.9586	0.01371
Peri-; endo-; and myocarditis; cardiomyopathy (except that caused by tuberculosis or sexually transmitted disease)	-0.039	0.96149	0.03121
Essential hypertension	-0.069	0.93335	0
Hypertension with complications and secondary hypertension	-0.081	0.92253	0.00026
Acute myocardial infarction	-0.049	0.95197	0.04293
Nonspecific chest pain	0.004	1.00346	0.83348
Pulmonary heart disease	0.046	1.04748	0.01013
Conduction disorders	-0.054	0.94785	0.00272
Cardiac arrest and ventricular fibrillation	-0.111	0.89513	0.03702
Congestive heart failure; non-hypertensive	-0.023	0.97723	0.12219
Acute cerebrovascular disease	-0.105	0.9003	0.00094
Transient cerebral ischemia	-0.067	0.93567	0.14652
Late effects of cerebrovascular disease	-0.169	0.84444	0
Peripheral and visceral atherosclerosis	-0.004	0.99635	0.8192
Hemorrhoids	0.014	1.01419	0.6243
Pneumonia (except that caused by tuberculosis or sexually transmitted disease)	-0.049	0.95218	0.00617
Influenza	-0.075	0.92738	0.14362
Acute bronchitis	-0.052	0.94953	0.05424
Other upper respiratory infections	-0.050	0.95093	0.02503
Chronic obstructive pulmonary disease and bronchiectasis	0.092	1.09627	0
Asthma	-0.019	0.9817	0.32031
Pleurisy; pneumothorax; pulmonary collapse	-0.029	0.97144	0.14122
Respiratory failure; insufficiency; arrest (adult)	0.021	1.02087	0.24263
Other lower respiratory disease	-0.028	0.9727	0.05377
Other upper respiratory disease	-0.069	0.93351	0.00275
Intestinal infection	-0.053	0.94856	0.04478

Covariate	Estimate	Odds Ratio	P-value
Diseases of mouth; excluding dental	0.145	1.15638	0.01049
Esophageal disorders	0.022	1.02268	0.10905
Gastroduodenal ulcer (except hemorrhage)	-0.024	0.97627	0.39495
Gastritis and duodenitis	-0.031	0.96906	0.1803
Other disorders of stomach and duodenum	0.074	1.07675	0.00123
Abdominal hernia	-0.051	0.95043	0.02548
Regional enteritis and ulcerative colitis	0.160	1.17296	0.01012
Intestinal obstruction without hernia	0.065	1.06682	0.07081
Diverticulosis and diverticulitis	-0.026	0.9739	0.28663
Peritonitis and intestinal abscess	-0.081	0.9221	0.01273
Other liver diseases	-0.039	0.96205	0.03022
Gastrointestinal hemorrhage	-0.024	0.97649	0.29637
Noninfectious gastroenteritis	-0.049	0.9522	0.07406
Nephritis; nephrosis; renal sclerosis	0.028	1.02841	0.29709
Acute and unspecified renal failure	0.021	1.02166	0.18122
Chronic kidney disease	0.012	1.0122	0.75177
Urinary tract infections	-0.008	0.99199	0.65902
Other diseases of kidney and ureters	-0.073	0.92946	0.0002
Genitourinary symptoms and ill-defined conditions	0.003	1.00341	0.8497
Other male genital disorders	0.066	1.06806	0.07377
Nonmalignant breast conditions	-0.035	0.96603	0.32803
Menopausal disorders	-0.025	0.97548	0.69471
Contraceptive and procreative management	0.010	1.00994	0.84579
Skin and subcutaneous tissue infections	0.006	1.00579	0.7407
Other inflammatory condition of skin	0.077	1.07992	0.00011
Chronic ulcer of skin	0.109	1.11525	0
Other skin disorders	-0.069	0.93361	0.00032
Infective arthritis and osteomyelitis (except that caused by tuberculosis or sexually transmitted disease)	0.045	1.04645	0.08991
Rheumatoid arthritis and related disease	0.189	1.20797	0
Osteoarthritis	0.100	1.10566	0
Other non-traumatic joint disorders	0.017	1.01687	0.30294
Spondylosis; intervertebral disc disorders; other back problems	0.194	1.21401	0
Osteoporosis	0.051	1.05267	0.08971
Acquired foot deformities	-0.066	0.93583	0.15961
Other acquired deformities	0.060	1.06176	0.08927
Other connective tissue disease	-0.005	0.99495	0.72989
Other bone disease and musculoskeletal deformities	-0.039	0.96227	0.03309
Digestive congenital anomalies	0.307	1.35905	0.00016
Genitourinary congenital anomalies	0.100	1.10555	0.00176

Covariate	Estimate	Odds Ratio	P-value
Nervous system congenital anomalies	-0.269	0.76398	0.02745
Joint disorders and dislocations; trauma-related	0.083	1.08673	0.16889
Spinal cord injury	-0.223	0.80004	0.1479
Fracture of upper limb	0.035	1.03587	0.38149
Fracture of lower limb	0.034	1.03419	0.31878
Sprains and strains	-0.065	0.93683	0.02243
Intracranial injury	-0.147	0.86358	0.00662
Open wounds of head; neck; and trunk	0.081	1.08445	0.04179
Open wounds of extremities	0.024	1.02409	0.3539
Poisoning by psychotropic agents	0.374	1.45412	0.00872
Poisoning by other medications and drugs	0.087	1.09108	0.06452
Poisoning by nonmedicinal substances	-0.094	0.91016	0.15689
Other injuries and conditions due to external causes	-0.003	0.9975	0.88747
Syncope	-0.098	0.9063	0.0002
Fever of unknown origin	0.026	1.02587	0.24395
Nausea and vomiting	0.043	1.04404	0.0074
Abdominal pain	-0.023	0.97721	0.18987
Malaise and fatigue	-0.072	0.93078	0.00002
Allergic reactions	0.042	1.04234	0.00387
Rehabilitation care; fitting of prostheses; and adjustment of devices	0.045	1.04616	0.52413
Administrative/social admission	-0.033	0.96726	0.09936
Medical examination/evaluation	-0.039	0.96185	0.02042
Other screening for suspected conditions (not mental disorders or infectious disease)	-0.020	0.98058	0.1758
Residual codes; unclassified	0.026	1.02622	0.09047
Adjustment disorders	0.016	1.01592	0.72919
Anxiety disorders	0.715	2.04336	0
Delirium dementia and amnestic and other cognitive disorders	-0.115	0.89114	0.00002
Developmental disorders	-0.147	0.86296	0.06367
Disorders usually diagnosed in infancy childhood or adolescence	0.350	1.41911	0.06508
Mood disorders	0.101	1.10613	0
Schizophrenia and other psychotic disorders	-0.018	0.98249	0.70663
Alcohol-related disorders	-0.109	0.89689	0.00166
Substance-related disorders	0.186	1.20437	0
Screening and history of mental health and substance abuse codes	0.046	1.04719	0.00071
Miscellaneous mental health disorders	0.142	1.1522	0.00116
External cause codes: Motor vehicle traffic (MVT)	0.043	1.04375	0.39326
External cause codes: Natural/environment	-0.054	0.94737	0.02303
External cause codes: Overexertion	0.175	1.1908	0.33187
Adverse effects of medical care	0.005	1.00535	0.70922

*cell intentionally left blank

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.**

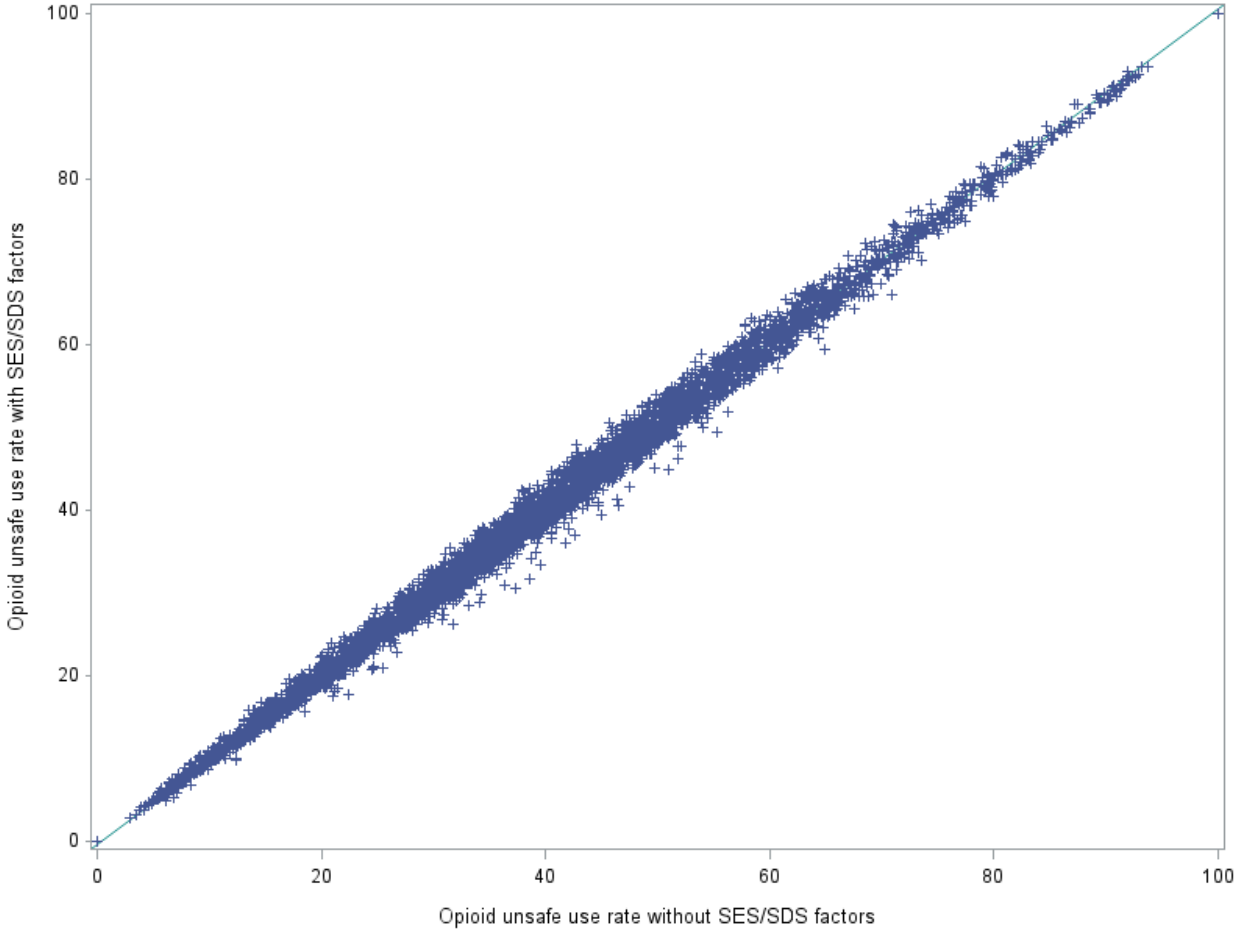
We fit an additional model including covariates from the original model and adding several SES/SDS indicators (dual-eligible insurance status, employment status at ESRD incidence, area deprivation index) as well as patient sex, race and ethnicity. Table 5 shows the associations from these selected additional covariates in the SES/SDS adjusted model.

Table 5. Coefficients and odds ratios for SDS/SES variables

Variable	Estimate	Odds Ratio	P-value
Sex	*	*	*
Female	-0.031	0.969	0.015
Male	Reference	*	0.000
Ethnicity	*	*	*
Hispanic	-0.305	0.737	<.0001
Non-Hispanic	Reference	*	0.000
Race	*	*	*
White	Reference	*	0.000
Black	-0.307	0.735	<.0001
Other	-0.522	0.594	<.0001
Employment Status (2728)	*	*	*
Employed	Reference	*	0.000
Unemployed	0.082	1.085	<.0001
Other	0.120	1.128	<.0001
Medicare Coverage	*	*	*
Dual eligible	0.072	1.075	<.0001
Non dual eligible	Reference	0.000	0.000
ADI (zipcode_level)	*	*	*
National percentile ADI score	-0.001	0.999	0.044

*cell intentionally left blank

Figure 3. Correlation between Opioid unsafe measures with and without SDS/SES adjustment



The correlation coefficient is 0.997 (p<.001).

Table 6. Comparison of flagging performances with .and without adjusting for SDS/SES factors

Measures	Opioid Unsafe use with SES/SDS: Better than Expected	Opioid Unsafe use with SES/SDS: As Expected	Opioid Unsafe use with SES/SDS: Worse than Expected	Opioid Unsafe use with SES/SDS: Total
Better than Expected	299 (5.8%)	10 (0.2%)	0	309 (6.3%)
As Expected	19 (0.4%)	4604 (89.9%)	12 (0.2%)	4635 (90.5%)
Worse than Expected	0	8 (0.2%)	171 (3.3%)	179 (3.5%)
Total	318 (6.2%)	4622 (90.2%)	183 (3.6%)	5123

These results show that there was only a small difference in the overall flagging rates between the models with and without SES/SDS adjusters. Specifically, fewer than 50 facilities (<1%) moved down or up one category and no facilities moved more than one category.

Though unsafe opioid use might be associated with white race, non-Hispanic ethnicity, dual eligible status, and unemployment (Table 5), these SDS/SES factors are not included in the final risk adjusted model as they play little roles in flagging; see the comments below Table 6. Furthermore, Figure 3 shows that unsafe opioid use measures based on models with and without SDS/SES factors are highly correlated. More importantly, while other studies have reported associations between patient-level race, ethnicity, and dual eligible status and unsafe opioid prescriptions, it is unclear whether these differences are due to underlying biological or other patient factors or represent disparities in care. Adjusting for these social risk factors could have the unintended consequence of creating or reinforcing disparities and facilitating unsafe prescribing practices. The primary goal should be to implement quality measures that result in the highest quality and safest patient care for all patients.

Finally, we comment that sex is the only SDS/SES factor that we include in our final risk adjustment model. Biologic differences (e.g., genetic, hormonal, metabolic) may account for differences in pain perception, suggesting a physiologic effect rather than a disparity in care.

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)
Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

Risk factors were selected for the final model based on the magnitude of the coefficients, evaluation of their statistical significance, and the model C-statistic. The C-statistic measures the discriminative power of the regression model with considered risk factors. Two-way interactions were examined and selected for the final model based on both the magnitude and statistical significance of the estimates.

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
The C-statistic (also known as the Index of Concordance) was 0.74, meaning that unsafe users had a 74% probability of having a higher model-based risk score than safe users and that the model has a good distinguishing power to distinguish unsafe users from safe users.

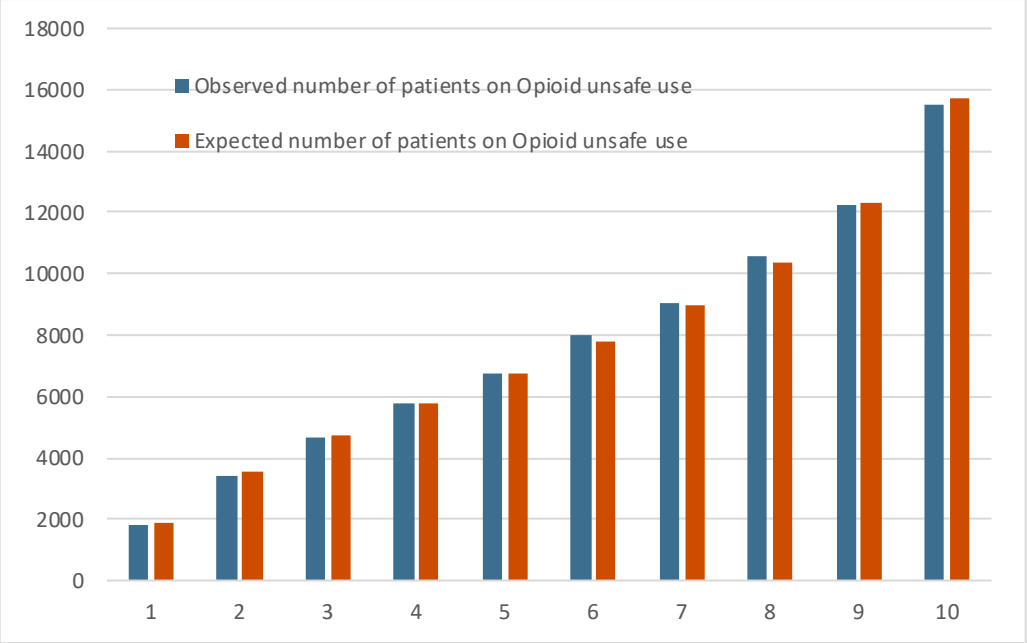
2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
To evaluate the goodness of fit of the developed risk prediction model, we ranked patients based on the predicted risk of unsafe use of opioids and grouped them into ten decile groups, with the lowest decile indicating the least risk and the highest decile the largest risk. Within each decile group, we computed the observed and expected numbers of unsafe users, based on our fitted model, as well as the standardized difference, i.e. (Obs-Exp)/Exp. It appears that the standardized difference was almost negligible (with the maximum absolute value being 0.0407), indicating a reasonably good fit of our model on the data.

Table 7. Observed and expected numbers of Opioid unsafe uses and their standardized differences within each decile group.

Observed number of patients on Opioid unsafe use	Expected number of patients on Opioid unsafe use	(Obs-Exp)/Exp
1779	1854.55	-0.0407
3423	3515.56	-0.0263
4646	4700.2	-0.0115
5754	5754	-0.0000
6774	6755.48	0.0027
7970	7790.6	0.0230
9039	8945.99	0.0104
10550	10394.91	0.0149
12272	12331.84	-0.0049
15520	15691.88	-0.0110

2b3.8. Statistical Risk Model Calibration: comparisons between expected and observed–:
Figure 4: Comparisons of expected and observed numbers of opioid unsafe users across decile groups (as defined in 2b3.7)

Figure 4. Observed and expected numbers of Opioid unsafe uses and their standardized differences within each decile group.



2b3.9. Results of Risk Stratification Analysis:

N/A

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

The decile plot shows that the risk factors in the model are discriminating well between patients. There is good separation among all 10 groups and the ordering is as predicted by the model (patients predicted to be at lower risk have less unsafe opioid use). The absolute differences between the groups are also large with patients predicted to have the highest unsafe opioid use (group 10) having a 4 times higher rate than those predicted to have the lowest unsafe use (group 1).

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

N/A

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

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

Differences in measure performance were evaluated separately for each prescriber group using patient level analyses. For each prescriber group, the proportion of patient-months with an unsafe opioid prescription, calculated at the year-level, was compared to the overall national distribution.

Note that the monthly based measure is a simple average of binary outcomes across individuals with the prescriber group, for which the binary outcome equals 0 if an unsafe opioid prescription is present and equals 1 if an unsafe opioid prescription is present. The differences in proportions can be compared using Fisher's Exact tests or its normal approximation. The yearly based measure, however, is not a simple average of binary outcomes and we instead used a re-sampling based exact test, with re-sampling generated from the population distribution of the patient level outcomes. To address the issue of over and under-dispersion of the data, we used the empirical null approach to flag facilities. More specifically, we first calculate the p-value by assessing the probability that patients with each dialysis practitioner group would experience a number of events more extreme than what was actually observed if the null hypothesis were true, where the null hypothesis is that a patient with each dialysis practitioner group will follow the overall national distribution. We then convert these p-values to z-scores. Using the mid-range of these scores (e.g., from the first quartile to the third quartile), we estimate the null distribution of the z-scores, which is termed the empirical null distribution. Finally, we use the 2.5 and 97.5 percentiles of the empirical null, as cut-offs, to determine the providers that fall into the categories of "worth than expected", "as expected" and "better than expected", respectively.

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*)

Table 8. Proportion of prescriber groups with statistically significant differences (p-value < 0.025)

Measures	Number of prescriber groups	Percent of prescriber groups
Better than Expected	309	6.03
As expected	4635	90.47
Worse than expected	179	3.49

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 annual percentage of patients with an unsafe opioid prescription as the performance measure, 4635 (90.5%) prescriber groups have achieved expected performance, 179 (3.5%) prescriber groups have performed worse than expected, and 309 (6.0%) have better than expected.

In general, lower rates of unsafe opioid prescriptions represent better quality of care. This analysis demonstrates both practical and statistically significant differences in performance across prescriber groups based on their proportion of patient months with unsafe opioid prescriptions.

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

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

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

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

N/A

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

N/A

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

what are the norms for the test conducted)
N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
Many data elements can be obtained from multiple sources and missing data occur rarely for covariates included in this measure. We assessed missing data for BMI which comes from form CMS 2728 which is the source of data used for the BMI risk adjustment in the model.

Ascertainment of prevalent comorbidities for risk adjustment relies on determining sufficient Medicare claims history. This is determined by the presence of 6 or more months of Medicare coverage in the prior 12 months OR 1 or more Medicare Advantage patient months in the prior 12 months. Medicare Advantage patient status was defined using Medicare Enrollment Database (EDB) criteria. We assessed the extent of incomplete comorbidity ascertainment for comorbidity risk adjustment.

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*)

Table 9. Frequency of missing data elements, 2017 data

Data Element	Missing (%)
Patients with missing CMS 2728	1.2%
Patients without BMI reported on 2728	2.1%
Patients where we are unable to determine presence of prevalent comorbidities	8.4%

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*)

There is a very low fraction of patients with missing BMI, missing form 2728, and missing cause of ESRD. Missing Cause of ESRD and missing 2728 were accounted for with a category for missingness in the model. Patients with missing BMI were included in the BMI 30+ category.

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

If other:

3b. Electronic Sources

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

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

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

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

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

Attachment:

3c. Data Collection Strategy

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

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

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

N/A

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).*

N/A

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting Payment Program	*

*cell intentionally left blank

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

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

N/A

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?*)

The measure is undergoing initial endorsement review.

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.*)

CMS will determine if/when to report this measure in a public reporting/payment program. One potential application for the measure is in the Quality Payment Program where it would be one of several optional measures that a group practice could select in their self-evaluation.

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

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

N/A

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

N/A

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

Describe how feedback was obtained.

N/A

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

N/A

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

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

N/A

Improvement

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

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

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

The measure is not yet implemented in a public reporting program, so improvement could not be evaluated. CMS currently anticipates implementation of this unsafe opioid measure. Once implemented prescriber performance on the measure can be evaluated to determine if the measure has supported and detected quality improvement in reducing unsafe opioid use, while accounting for patients where higher dose or longer term therapy may be warranted.

4b2. Unintended Consequences

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

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

N/A

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

N/A

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

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

Are the measure specifications harmonized to the extent possible?

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

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.)

Appendix

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

Attachment Attachment: Prescriber_Appendix.pdf

Contact Information

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

Co.2 Point of Contact: Kimberly, Rawlings, Kimberly.Rawlings@cms.hhs.gov

Co.3 Measure Developer if different from Measure Steward: University of Michigan Kidney Epidemiology and Cost Center

Co.4 Point of Contact: Jennifer, Sardone, jmsto@med.umich.edu

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

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

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2021

Ad.3 Month and Year of most recent revision: 04, 2021

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

Ad.5 When is the next scheduled review/update for this measure? 04, 2022

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: